



EDITED BY

EYAL

ZAMIR

DORON

TEICHMAN

≡ The Oxford Handbook *of*  
BEHAVIORAL ECONOMICS  
AND THE LAW

THE OXFORD HANDBOOK OF

**BEHAVIORAL  
ECONOMICS AND  
THE LAW**

## CONSULTING EDITORS

Michael Szenberg  
*Lubin School of Business, Pace University*

Lall Ramrattan  
*University of California, Berkeley Extension*

THE OXFORD HANDBOOK OF

---

**BEHAVIORAL  
ECONOMICS  
AND THE LAW**

---

*Edited by*  
EYAL ZAMIR  
*and*  
DORON TEICHMAN

**OXFORD**  
UNIVERSITY PRESS

OXFORD  
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.  
It furthers the University's objective of excellence in research, scholarship,  
and education by publishing worldwide.

Oxford New York  
Auckland Cape Town Dar es Salaam Hong Kong Karachi  
Kuala Lumpur Madrid Melbourne Mexico City Nairobi  
New Delhi Shanghai Taipei Toronto

With offices in  
Argentina Austria Brazil Chile Czech Republic France Greece  
Guatemala Hungary Italy Japan Poland Portugal Singapore  
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trademark of Oxford University Press  
in the UK and certain other countries.

Published in the United States of America by  
Oxford University Press  
198 Madison Avenue, New York, NY 10016

© Oxford University Press 2014

All rights reserved. No part of this publication may be reproduced, stored in a  
retrieval system, or transmitted, in any form or by any means, without the prior  
permission in writing of Oxford University Press, or as expressly permitted by law,  
by license, or under terms agreed with the appropriate reproduction rights organization.  
Inquiries concerning reproduction outside the scope of the above should be sent to the  
Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form  
and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data  
The Oxford handbook of behavioral economics and the law/edited by  
Eyal Zamir and Doron Teichman.  
pages cm. — (Oxford handbooks)  
Includes bibliographical references and index.  
ISBN 978-0-19-994547-4 (alk. paper)  
1. Economics—Psychological aspects—Handbooks, manuals, etc.  
2. Law and economics—Handbooks, manuals, etc.  
I. Zamir, Eyal, editor of compilation.  
II. Teichman, Doron, editor of compilation.  
HB74.P8O94 2014  
330.01'9—dc23  
2014006783

1 3 5 7 9 8 6 4 2  
Printed in the United States of America  
on acid-free paper

# CONTENTS

---

<i>Contributors</i>	ix
<i>Introduction</i>	xi

## PART I. BEHAVIORAL ECONOMICS: AN OVERVIEW

1. Heuristics and Biases	3
JONATHAN BARON	
2. Human Prosocial Motivation and the Maintenance of Social Order	28
SIMON GÄCHTER	
3. Moral Judgment	61
JONATHAN BARON	

## PART II. BEHAVIORAL ECONOMICS AND THE LAW: AN OVERVIEW AND CRITIQUE

4. The Importance of Behavioral Law	93
THOMAS S. ULEN	
5. Behavioral Law and Economics: Empirical Methods	125
CHRISTOPH ENGEL	
6. Biasing, Debiasing, and the Law	143
DANIEL PI, FRANCESCO PARISI, AND BARBARA LUPPI	
7. Alternative Behavioral Law and Economics	167
GREGORY MITCHELL	

### **PART III. BEHAVIORAL ECONOMICS AND THE LAW: SPECIFIC BEHAVIORAL PHENOMENA**

- 8. Law and Prosocial Behavior 195  
LYNN A. STOUT
- 9. Behavioral Ethics Meets Behavioral Law and Economics 213  
YUVAL FELDMAN
- 10. Law, Moral Attitudes, and Behavioral Change 241  
KENWORTHY BILZ AND JANICE NADLER
- 11. Law's Loss Aversion 268  
EYAL ZAMIR
- 12. Wrestling with the Endowment Effect, or How to Do Law and  
Economics without the Coase Theorem 300  
RUSSELL KOROBKIN
- 13. Probability Errors: Overoptimism, Ambiguity Aversion, and the  
Certainty Effect 335  
SEAN HANNON WILLIAMS
- 14. The Hindsight Bias and the Law in Hindsight 354  
DORON TEICHMAN

### **PART IV. BEHAVIORAL ECONOMICS: LEGAL APPLICATIONS**

- 15. Behavioral Law and Economics of Property Law: Achievements and  
Challenges 377  
DAPHNA LEWINSOHN-ZAMIR
- 16. Behavioral Economics and Tort Law 405  
YOED HALBERSBERG AND EHUD GUTTEL
- 17. Behavioral Economics and Contract Law 438  
MELVIN A. EISENBERG
- 18. Consumer Transactions 465  
OREN BAR-GILL



---

19. Behavioral Economics and Insurance Law: The Importance of Equilibrium Analysis	491
TOM BAKER AND PETER SIEGELMAN	
20. The End of Contractarianism? Behavioral Economics and the Law of Corporations	518
KENT GREENFIELD	
21. The Market, the Firm, and Behavioral Antitrust	539
AVISHALOM TOR	
22. Behavioral Analysis of Criminal Law: A Survey	568
ALON HAREL	
23. Behavioral Economics and the Law: Tax	599
EDWARD J. McCAFFERY	
24. Litigation and Settlement	623
JENNIFER K. ROBBENOLT	
25. Behavioral Economics and Plea Bargaining	643
RUSSELL COVEY	
26. Judicial Decision-Making: A Behavioral Perspective	664
DORON TEICHMAN AND EYAL ZAMIR	
27. Evidence Law	703
FREDRICK E. VARS	
28. Nudges.gov: Behaviorally Informed Regulation	719
CASS R. SUNSTEIN	
29. Environmental Law	748
ADRIAN KUENZLER AND DOUGLAS A. KYSAR	
<i>Index of Names</i>	783
<i>Subject Index</i>	807





## CONTRIBUTORS

---

**Tom Baker** is William Maul Measey Professor of Law and Health Sciences at the University of Pennsylvania Law School.

**Oren Bar-Gill** is Professor of Law and Economics at Harvard Law School.

**Jonathan Baron** is Professor of Psychology at the University of Pennsylvania.

**Kenworthy Bilz** is Professor at the University of Illinois College of Law.

**Russell Covey** is Professor of Law at Georgia State University's College of Law.

**Melvin A. Eisenberg** is Jesse H. Choper Professor of Law at the University of California-Berkeley School of Law.

**Christoph Engel** is Director at the Max Planck Institute for Research on Collective Goods.

**Yuval Feldman** is Associate Professor of Law at Bar-Ilan University and Senior Fellow at the Israel Democracy Institute.

**Simon Gächter** is Professor of the Psychology of Economic Decision Making in School of Economics at the University of Nottingham.

**Kent Greenfield** is Professor of Law and Dean's Research Scholar at Boston College Law School.

**Ehud Guttel** is Frieda & Solomon B. Rosenzweig Chair in Law at the Hebrew University of Jerusalem.

**Yoed Halbersberg** is Adjunct Professor of Law at the Hebrew University of Jerusalem, and Research Fellow and Adjunct Professor of Law at Bar Ilan University.

**Alon Harel** is Phillip P. Mizock & Estelle Mizock Chair in Administrative and Criminal Law at the Hebrew University of Jerusalem.

**Russell Korobkin** is Richard C. Maxwell Professor of Law at the UCLA School of Law.

**Adrian Kuenzler** is Branco Weiss Fellow of Society in Science at Yale Law School.

**Douglas A. Kysar** is Joseph M. Field '55 Professor of Law at Yale Law School.

**Daphna Lewinsohn-Zamir** is Louis Marshall Professor of Environmental Law in the Faculty of Law at the Hebrew University of Jerusalem.

**Barbara Luppi** is Assistant Professor in Political Economics at the University of Modena and Reggio Emilia and Visiting Professor at the University of St. Thomas School of Law.

**Edward J. McCaffery** is Robert C. Packard Trustee Chair in Law, Economics, and Political Science at the University of Southern California Gould School of Law and Visiting Professor of Law and Economics at the California Institute of Technology.

**Gregory Mitchell** is Joseph Weintraub–Bank of America Distinguished Professor of Law at the University of Virginia School of Law.

**Janice Nadler** is Research Professor at the American Bar Foundation and Professor at Northwestern University School of Law.

**Francesco Parisi** is Oppenheimer Wolff and Donnelly Professor of Law at the University of Minnesota Law School and Professor of Economics at the University of Bologna.

**Daniel Pi** is an Erasmus Mundus Fellow and Ph.D. candidate in the European Doctorate in Law and Economics at the University of Bologna, University of Hamburg, and Erasmus University Rotterdam.

**Jennifer K. Robbennolt** is Professor of Law and Psychology at the University of Illinois College of Law.

**Peter Siegelman** is Roger Sherman Professor of Law at the University of Connecticut School of Law.

**Lynn A. Stout** is Distinguished Professor of Corporate and Business Law at Cornell Law School.

**Cass R. Sunstein** is Robert Walmsley University Professor at Harvard Law School.

**Doron Teichman** is Joseph H. and Belle R. Braun Associate Professor in the Faculty of Law at the Hebrew University of Jerusalem

**Avishalom Tor** is Professor of Law and Director of the Research Program on Law and Market Behavior at the University of Notre Dame Law School.

**Fredrick E. Vars** is Associate Professor of Law at the University of Alabama School of Law.

**Thomas S. Ulen** is Swanlund Chair Emeritus at University of Illinois at Urbana-Champaign and Professor Emeritus of Law at University of Illinois College of Law.

**Sean Hannon Williams** is Professor at the University of Texas School of Law.

**Eyal Zamir** is Augusto Levi Professor of Commercial Law at the Hebrew University of Jerusalem.

# INTRODUCTION

---

For several decades, one of the leading perspectives in legal theory—perhaps *the* leading perspective—has been the economic analysis of law. The theory of human behavior underlying standard economic analysis of law, like economic analysis in general, has been the rational choice theory. According to this theory, people strive to enhance their own well-being, choosing the available option that would maximize their expected utility. In the past two decades or so, hand in hand with comparable developments in economics, the economic analysis of law has been challenged by a growing body of experimental and empirical studies attesting to prevalent and systematic deviations from the assumptions of economic rationality. These studies have challenged the assumption of thin, cognitive rationality by showing that people’s preferences often do not comply with the rules of dominance, transitivity, and invariance. These studies also called into question the assumption of thick, motivational rationality by pointing to the role of motivations such as envy and altruism. From a slightly different angle, experimental and empirical studies have shown that people’s moral judgments do not fall in line with the consequentialist underpinnings of welfare economics, the normative branch of economic analysis, but are much more aligned with deontological morality.

Initially perceived as an antithesis to standard law and economics, over time these insights have been largely integrated into mainstream economic analysis of law. As Russell Korobkin (2012) noted, “the battle to separate the economic analysis of legal rules and institutions from the straightjacket of strict rational choice assumptions has been won.”<sup>1</sup>

The introduction of behavioral analysis has not only had a profound effect on the economic analysis of law. Arguably, the behavioral movement has been one of the most influential developments in legal scholarship in recent years.<sup>2</sup> To a large degree, much in the way that economic reasoning became a standard form of legal analysis in the 1980s and 1990s, behavioral analysis became a standard form of analysis in the past decade or so. Behavioral insights possibly have a greater impact on legal analysis

<sup>1</sup> Thus, for example, whereas in the past behavioral studies presented at the American Law and Economics Association annual meeting were delegated to separate panels, they are currently an integral and major part of the program.

<sup>2</sup> As anecdotal evidence of this point it is interesting to observe that Jolls, Sunstein, and Thaler (1998) is the most cited legal article published in the last twenty years (even when compared to articles that were published seven years ahead of it) (Shapiro and Pearse 2012).

than on other spheres of economic analysis, as jurists are often particularly interested in real-life policymaking, which must take into account people's actual behavioral characteristics.

In recent years, the growing influence of behavioral law and economics has been accompanied by the emergence of empirical and experimental legal studies. This new paradigm has transformed the nature and scope of the research conducted by behavioral-legal scholars. Rather than just draw on the results of experimental and empirical studies in the social sciences, a steadily growing number of researchers engage in experimental and empirical studies designed specifically to answer questions that are of especial interest to jurists.

Thanks to these developments, the integration of economics, psychology, and law is breaking exciting new ground in legal theory and the social sciences. Despite the burgeoning academic interest in behavioral law and economics, to date no book offers a general introduction that presents the state of the art of the field.<sup>3</sup> This volume contains new contributions that together provide an introductory treatment and critical overview of behavioral economics and the law. It aims to serve both as a gateway into behavioral law and economics and as a catalyst for new scholarly research.

The *Handbook* consists of twenty-nine chapters organized in four parts. The first part provides a general overview of behavioral economics. It consists of three chapters, dealing with heuristics and biases, human motivation, and moral judgments, respectively. The second part comprises four chapters introducing and criticizing the contribution of behavioral economics to legal theory. Chapter 4 is a general introduction to the field of behavioral law and economics; chapter 5 discusses the empirical methodologies used in behavioral legal studies; chapter 6 analyzes different techniques of debiasing and insulation through the law; and chapter 7 challenges some of the characteristics of current research in behavioral law and economics.

The third part of the *Handbook* discusses specific behavioral phenomena, their ramifications for legal policymaking, and their reflection in extant law. Some of the seven chapters in this part refer to general phenomena, others to more specific ones. These include prosocial behavior, bounded ethicality, the inculcation of moral attitudes, loss aversion, the endowment effect, probability errors, and the hindsight bias (chapters 8 to 14, respectively). Finally, the fourth part—chapters 15 through 29—analyzes the contribution of behavioral economics to fifteen legal spheres. These are property law, tort law, contract law, consumer transactions, insurance law, law of corporation, antitrust law, criminal law, tax law, litigation and settlement, plea bargaining, judicial decision-making, evidence law, regulation, and environmental law.

<sup>3</sup> Notable anthologies of previously published studies are Sunstein (2000) and Rachlinski (2005).

**REFERENCES**

---

- Jolls, C., C. R. Sunstein, and R. Thaler. 1998. A Behavioral Approach to Law and Economics. *Stanford Law Review* 50:1471–550.
- Korobkin, R. 2011. What Comes after Victory for Behavioral Law and Economics. *University of Illinois Law Review* 2011:1653–74.
- Rachlinski, J. J. (editor). 2009. *Behavioral Law and Economics*. Cheltenham Glos: Edward Elgar Publishing.
- Shapiro, F. R., and M. Pearse. 2012. The Most Cited Law Review Articles of All Times. *Michigan Law Review* 110:1483–520.
- Sunstein, C. R. (editor) 2000. *Behavioral Law and Economics*. Cambridge: Cambridge University Press.





PART I

---

**BEHAVIORAL  
ECONOMICS**

*An Overview*

---



## CHAPTER 1

---

# HEURISTICS AND BIASES

---

JONATHAN BARON

THE field I shall characterize as Judgment and Decision Making (JDM) approaches research questions from three perspectives: normative, prescriptive, and descriptive. These have been implicit in various writers for a long time, but the current field began as a reaction to work in cognitive psychology in the 1950s and 1960s. Around that time, cognitive psychologists, as they came to be called later, were studying performance in mental tasks of various sorts, including estimation of probability of red balls drawn from urns, judgments of statistical properties of sequences of numbers, and decisions concerning simple gambles. (For reviews, see Peterson & Beach 1967, and Davidson, Suppes, & Siegel 1957.) At the time, various idealized models of how tasks should be done had been developed, such as information theory, expected-utility theory, Bayesian probability theory, statistics, and (a little later) signal-detection theory. Psychologists were asking whether these (mostly mathematical) models described actual performance. At first the answer seemed to be yes, but results questioning these models began to accumulate to the breaking point.

If there was an actual point at which the break occurred, it was after Ward Edwards and his students had tried to salvage Bayesian probability theory as an account of human probability judgment—of the probability of drawing a red ball from an urn—after experience with the urn, by adding the assumption that people were Bayesian but conservative, less willing to change their beliefs than the theory said they should be but otherwise consistent with it. In 1972, reporting some experiments, Kahneman and Tversky wrote, “In his evaluation of the evidence, man is apparently not a conservative Bayesian: he is not Bayesian at all” (p. 450). Instead, Kahneman and Tversky (henceforth K&T) proposed that people make probability judgments by using a heuristic, a rule of thumb, which they called *representativeness*. The Bayesian theory said that people should start out with some prior opinions about the composition of the urn, then modify those opinions systematically as the evidence came in from repeated draws. K&T proposed that, instead, people examine the evidence and judge its similarity to various possible proportions.

There are interesting complexities in this sequence of experiments, but the important thing was that K&T had now made a distinction between the normative theory, which describes the judgments that should be made, and what people actually do. People are systematically biased in these cases. As other work showed, people ignore information about base rates, that is, about what should play the role of their prior beliefs in the Bayesian theory. And the reason they are biased is that they use a simple rule, a heuristic, one that will approximate the right answer under some conditions, and one that is much easier to use than a principle that requires taking two quantities into account, which could pull in different directions. Thus was born the idea of heuristics and biases. The theory that people use a heuristic was a descriptive model, while Bayesian theory remained the normative model, the standard for what would count as the best answer.

The field of economics took longer to recognize a distinction between normative and descriptive theory. Many economists still argue that decisions are approximately rational and that our task is to discover their hidden rationality rather than to look for departures from normative models. But the psychological work on biases has had substantial influence on economics, especially experimental economics, which generally assumes that the descriptive principles of decision-making are applicable, so that they may be studied under the carefully controlled conditions of the laboratory as well as in the real world.

The normative/descriptive distinction implied that people were systematically departing from optimal judgments and decisions. The next question for practical people is what to do about it, if anything can be done. That is the prescriptive part of the three-way distinction. In this chapter, I shall review the three types of theory, giving examples of each, with special attention to the variety of descriptive principles, but very little attention to prescriptive ideas.

## 1 NORMATIVE, DESCRIPTIVE, PRESCRIPTIVE

---

The three-way distinction emerged clearly in the 1980s (Freeling 1984; Baron 1985; Bell, Raiffa, & Tversky 1988—all of whom wrote independently of each other), although various parts of it were implicit in the writing of Herbert Simon and many philosophers (such as J. S. Mill).

Normative models, as noted, are standards for evaluation. They must be justified independently of observations of people's judgments and decisions, once we have observed enough to define what we are talking about. Sometimes they are obvious, as when we study people's judgment of length or duration. The normative model is the right answer, the objective length or duration of what people judge. When not obvious,

normative models are typically justified by philosophical and mathematical argument (Baron 2004).

Descriptive models are psychological theories that try to explain how people make judgments and decisions, typically in the language of cognitive psychology, which includes such concepts as heuristics and strategies, as well as formal mathematical models. Within the three-model framework, descriptive models are most useful when they explain departures from normative models, so researchers often focus on the search for such explanations. Such models allow us to determine whether, and, if so, how, we might improve judgments and decisions. When a deviation from a normative model is found to be systematic, not just the result of random error, we call it a bias. For example, people are biased to choose default options, even when others are normatively equal or better.

Prescriptive models are designs for improvement. If normative models fall in the domain of philosophy (broadly defined) and descriptive models in the domain of empirical psychological science, then prescriptive models are in the domain of design, engineering, or practice (again, broadly defined), roughly the sorts of things that are taught in professional schools rather than schools of “arts and sciences.” Originally, prescriptive models were conceived as including mathematical tools that were useful for the formal analysis of decisions. These constitute the field of decision analysis, which includes several methods (and which has a society and a journal by that name). But prescriptive models can also be educational interventions (Larrick 2004), which, for example, teach people alternative heuristics, to counteract heuristics that lead to biases.

A recent addition to the arsenal of prescriptive methods is the idea of “decision architecture” (Thaler & Sunstein 2008; chapter 28 by Sunstein in this volume), which consists of designing the presentation of decisions to those who will make them in such a way as to help people make the normatively better choice. A classic type of case is using the fact that people are biased toward the default to help them choose wisely by making what is usually the wise choice the default. For example, use a diversified portfolio as the default retirement plan for new employees (as opposed to, say, shares in company stock).

Thus, the ideal plan for JDM, sometimes actually realized (Baron 2008; Thaler & Sunstein 2008), is to apply normative models to judgments and decisions, looking for possible biases, then use the tools of psychology to understand the nature of those biases, and then, in the light of this understanding, develop approaches to improve matters. Of course, in real life these steps are not sequential, but are informed by each other. For example, formal decision analysis turns out to require the measurement of personal probability and utility, so now a large descriptive and normative enterprise is devoted to this measurement problem, which has produced better methods for measurement, which, in turn, are used to improve the original prescriptive models.

This plan clearly requires that the three elements are kept distinct. Suppose, for example, we make arguments for normative models on the basis of (descriptive) observations of what people do, under the assumption that people are rational. Then, we are

likely to conclude that people are rational and that no prescriptive interventions are needed. The field of JDM would tend to disappear. Arguably, economics as a field made this assumption of rationality and thus was never concerned with helping people to make better economic choices, until recently.

Another danger that JDM tries to avoid is to design prescriptive interventions without at least some clarity about normative and descriptive models. Specifically, we try to avoid “fixing things that ain’t broke.”

Much of the debate within JDM is about the seriousness of various purported biases. Although strong advocates on one side or the other tend to think either that people are hopelessly biased or that we are perfectly adapted to our environment, more moderate folks think that, while it all depends on the person, the situation, and the task, there really are some situations where people can be helped, sometimes a lot, through the JDM approach (Thaler & Sunstein 2008). Of course there are also many judgments and decisions that people make amazingly well. These extremely good judgments are mostly studied under the topic of expertise (e.g., Ericsson & Ward 2007). An example is the study of expert forecasting of political events, in which the determinants of successful forecasting, and the methods of good forecasters, are examined (Tetlock 2005). I will not discuss expertise in what follows.

We need to keep normative and prescriptive models separate as well. If we assume that normative models are also prescriptive, they may become self-defeating. In decision-making, the main normative standard is the maximization of (expected) utility, and the time required for calculation usually reduces utility. If normative models require elaborate calculation, then, when a real person attempts to apply one to a decision, the utility loss from the time spent may be greater than the gain from using the model, as opposed to some simpler heuristic. In many cases, then, normative models are applied by researchers, and real people may use various heuristics to improve their judgments as evaluated by the normative models (e.g., Davis-Stober, Dana, & Budescu 2010).

On the other hand, summary versions of normative models may require no calculation at all and may serve the purpose of focusing attention on only what is relevant (Baron 1990). For example, on a day in which it seemed that there was roughly an even chance that it would be cloudy or sunny, my wife said, “Should I wear sunglasses or regular glasses? I guess sunglasses because I would be more upset if it is sunny and I don’t wear them than if it is cloudy and I do.” This is an expected-utility calculation, but without any numbers.

## 1.1 Normative Theories

This section discusses the major normative theories. These theories cannot be derived from observation of people’s judgments, for then we would not be able to criticize them. Nor can we derive them from an apparent consensus of scholars, as the majority might

be wrong. Of course, any other way of developing normative models must be tentative, but it seems more fruitful to look at the actual arguments that scholars make, rather than just their vote on the question of which theories are correct. The argument that “scholars disagree” leads to no conclusion on its own, but it is frequently an excuse to avoid the issue.

So how should normative theories be developed? I have argued (2004) that they can be developed by imposing an analytic scheme on the world. The analytic scheme has to fit the question reasonably well, so it must be responsive to properties of human beings that lead them to ask normative questions in the first place. But it does not need to take into account their intuitive answers to the questions.

Of course, for some purposes, normative models are much simpler than this. If the task is for people to judge the populations of cities, then the normative model is simply the right answers.

### 1.1.1 *Utility Theory and Utilitarianism*

Utility theory says that a decision can be analyzed into options, each option can be assigned a number representing its goodness, that is, its utility, and the best option is the one with the highest utility. Variants of utility theory differ in how they analyze options. The main variants are the following:

- *Expected utility* (EU) assumes that the outcomes of options are uncertain, and the uncertainty results from states of the world, each of which has a probability. Each outcome has a utility, and the expected utility of an option is the result of multiplying each option’s utility by its probability and then adding up the results across the uncertain states.
- *Multiattribute utility* (in its simplest form) analyzes options into independent, nonoverlapping attributes, each of which has a weight. The weights replace the probabilities in EU. Total utility of an option is again a weighted sum of attribute utilities.
- In *discounted utility*, my name for something that lacks a name, the analysis is into times at which outcomes occur, rather than uncertain states or attributes. The usual result is that the weight is an exponential (decay) function of temporal delay.
- In *utilitarianism*, the units of analysis are people, rather than possible states of the world, attributes, or times.

Variants may be combined for a more complex analysis.

An example of the use of an analytic scheme is for EU. We analyze choices into options, states of the world, and outcomes. The difference between options and states is that the former are under the decision-maker’s control. The argument for EU as normative follows from additional assumptions about what the task is, and logic (Baron 2004, 2008).



### 1.1.2 *Probability and Statistics*

In JDM, probability is usually (but not always) understood as a measure of degree of belief.<sup>1</sup> When probability is defined this way, the basic mathematical principles can be derived from EU theory, by thinking about the logical properties of bets. The two main principles are the multiplication rule and the addition rule, which apply to probabilities as numbers you assign to your beliefs. The multiplication rule says that your probability for propositions A and B both being true is the product of your probability for B alone and your probability for A, conditional on B being true, which is notated as  $p(A|B)$ . The addition rule says that your probability for “A or B,” when you believe that A and B cannot both be true, is the sum of your probabilities for the two propositions. These two rules can be combined to yield Bayes’ theorem, which specifies the relation between  $p(A|B)$  and  $p(B|A)$ , that is, how to reverse conditional probabilities.<sup>2</sup>

We may test people’s probability judgments for related propositions by looking for violations of rules of *coherence* like these. A different normative model for probability is based on *correspondence* with reality. One measure of this is calibration, which is based on the idea that, if we look at the truth of all propositions where your probability is .75, then 75% of them should be true.

Researchers in JDM also use statistics as a normative model, for example, asking whether people apply the principle of regression to the mean in their own quantitative judgments.

Another much studied normative model is discounted utility, which concerns utilities of events that occur at different times. A simple normative model holds that time should not matter, just as a person’s identity should not matter for utilitarianism. You should think of yourself as a stream of different people sharing very similar memories and goals. But, just as EU theory for self-interested decisions can conflict with utilitarianism, so it can be argued that bias toward the present may be rational. Yet it turns out that, once we allow bias toward the present, we are still constrained by a the principle of dynamic consistency, which holds that the choice between two options occurring at two different times should not depend on when the choice is made (provided of course that nothing about the consequences of the two options changes with time). If people follow this principle, then the effective utility of an option may decline with time, but it must decline as an exponential function (i.e., in each unit of time, the utility declines by a constant proportion of its value at the end of the last unit). Interestingly, real decisions often weigh the immediate future much more than this model would require (Baron 2008).

<sup>1</sup> An alternative conception in which probability is a relative frequency can be subsumed under the topic of judgment: the question “How likely is a randomly selected person to be over six feet tall?” is similar enough to the question “What proportion of people are over six feet tall?”

<sup>2</sup> Specifically,  $p(B|A) = p(A|B)p(B)/p(A)$ .

## 1.2 Types of Descriptive Theory

Descriptive theories are drawn from the language of cognitive psychology, broadly conceived. They do not usually attempt to reduce judgments to some other level of explanation. That said, there is now a field of “decision neuroscience” and the field of evolutionary psychology sometimes deals with issues in JDM, another type of biological explanation. I shall leave these aside here.

Psychological explanation, in cognitive psychology, is mostly in terms of mental representations and processes. Representations can be perceptions, beliefs, emotions, and goals. (The inclusion of emotions and goals is why I define “cognitive” broadly.) In JDM, three kinds of representations are of particular interest (Baron 1985): possibilities, evidence, and goals. All thinking can be described in these terms. And thinking may also be divided into search and inference. Search is looking for these three objects, and inference is arriving at some conclusion (possibly tentative) from them, such as a judgment or choice. Much, but not all, explanation assumes that people are “intentional systems,” which means that we can understand them in terms of their goals and beliefs.

One type of process is a heuristic, a rule that a person applies, such as, “In a choice between two gambles in which the probabilities of winning are close, choose the one with the higher maximum payoff.” Note that such a rule could lead to violations of EU, because the difference in probabilities is completely ignored if it is below a certain threshold.

## 1.3 Mathematical Models

In JDM, a major concept is the idea of heuristics. Before turning to this idea in detail, I should also mention that JDM also relies heavily on mathematical models of processes or representations. Indeed, JDM has developed along with the field of mathematical psychology.

For example, a major research question has concerned how people combine representations of value and belief when making choices. EU theory says that the combination rule should be simple multiplication of utility (a measure of value) and probability (a measure of belief). Several theories (including the famous “prospect theory” of Kahneman and Tversky 1979) proposed that both quantities were transformed before multiplication, and other theories (e.g., the TAX model of Birnbaum 2008) have argued that other kinds of modifications of EU are required to account for choices in experiments on gambles. The predictions of models based on heuristics can be derived within the same mathematical framework. Heuristics predict certain kinds of deviations from EU, and from other models.

Another type of mathematical model concerns choices over time. As noted, discounted utility as a normative model implies that the effect of an outcome on choice, its decision utility, should decline exponentially as the outcome gets farther away in time

(or not decline at all, which is the limiting case of very slow decay). Researchers have found that other functions, such as a hyperbola, fit the data better, leading to the idea of “hyperbolic discounting” (Frederick, Loewenstein, & O’Donoghue 2002). This result is of major concern in economics, because it affects decisions involving money, such as investing, spending, and saving. If people behave as if they value the immediate future too much, they will save too little. More generally, such biases require self-control to overcome. They affect health, work, and personal relationships as well as financial status.

Yet another type of mathematical model describes hypothetical internal processes of choice or judgment. Some models of this type are based on the statistics of multiple regression and involve weights of various cues (e.g., Bröder & Newell 2008). Others involve processes that take place in time, so that they account for response times as well as choices (e.g., Glöckner 2009). For example, the process of arriving at a choice is conceived as a random walk, like a drunk who starts somewhere in the middle of a football field and sets out in some direction (e.g., Milosavljevic et al. 2010). If he walked straight, he would go out of bounds on the side toward which he started, but he cannot walk straight, so he sometimes goes out of bounds on the other side, although this will usually take longer.

I mention these models because it is often said that the field of JDM is dominated by the heuristics-and-biases (HB) approach and its critics, yet much of the field has little to do with heuristics, although it still involves biases, usually.

## 2 HEURISTICS AND THEIR HISTORY

---

But much of the field does have to do with heuristics. The idea has an interesting history.

### 2.1 Polya

The term “heuristic” was invented by George Polya (1945), a mathematician who was trying to explain how mathematicians think, with the goal of recommending these methods to students. A heuristic is distinguished from an algorithm, a rule with very clear conditions for its application, such as the method we learn in school for regrouping (borrowing) in subtraction, or the method for long division. Algorithms always do what they are supposed to do. By contrast, a heuristic is a rule without very clear conditions, and it does not always do anything useful. Sometimes it can suggest something else, which can subsequently lead to a solution. Examples are “Do you know a related problem?” or “Draw a figure” or “Could you derive something useful from the data?” The use of heuristics would give problem solvers something to do when they were stuck. Thus, heuristics were never harmful and possibly beneficial.

## 2.2 Kahneman and Tversky and the Classic Heuristics and Biases

Kahneman and Tversky (1972) appropriated the idea of heuristics to explain biases in probability judgment. They found that, in certain situations, people made judgments inconsistent with the rules of probability, particularly Bayes' theorem. They proposed that people were not attempting to apply these rules but rather were using a heuristic based on the similarity between a given case and a group as a way of judgment the probability that the case was a member of the group. This similarity, or representativeness, is relevant, but the use of only this cue ignored the base rates, the size of the group relative to alternative groups.

For example, Kahneman and Tversky (1973) gave subjects a personality sketch of a graduate student, Tom W, which was designed to match the stereotype of a student in computer science. They asked student subjects to rank various fields for the likelihood that Tom was a student in each field. These rankings were almost perfectly correlated with other students' ranking of similarity of the description to their stereotypes of each field (i.e., representativeness). Although students knew that computer science, at the time, was a small field, they completely ignored this knowledge in making their judgments. (In case the size isn't obviously relevant, use a Polya-type heuristic, which is to consider an extreme case, namely, the case in which the size of the field is 0.) Kahneman and Tversky thus concluded (from this and other evidence) that people were using a representativeness heuristic, and this led to biases away from the normative model, which would take into account base rates as well as similarity.

Another heuristic studied by Kahneman and Tversky (e.g., 1973) was the use of availability as a cue. Probability judgments were affected by the ease of thinking of examples and the number of examples thought of. In the case of availability, it is difficult to see what else people could do, without looking up the answer. This contrasts with representativeness, where people did have relevant knowledge but did not use it.

For Kahneman and Tversky, then, the idea of a heuristic was two-edged. Heuristics were useful because they saved time (and presumably allowed someone who had no idea of probability to answer the questions at all). And heuristics often gave useful answers, for example, when the sizes of the relevant groups were fairly close. But the idea of heuristics was invoked to explain biases, which were departures from normative models. The import of this HB approach was to explain why heuristics were functional and reasonable yet still led to errors in situations that were sometimes important and possibly even frequent. Some of the initial illustrations of heuristics involved neglect of base rates in the law and in medical diagnosis (Tversky & Kahneman 1982).

This approach to heuristics led to the postulation of a great many more heuristics by Kahneman, Tversky, and many other researchers. It has been, and continues to be, a fruitful approach to the study of judgments and decisions.

### 2.3 Gigerenzer's Critique: Heuristics Are Fast and Frugal

Gerd Gigerenzer and his colleagues, over many years (e.g., Gigerenzer, Todd, & the ABC Research Group 1999), criticized the heuristics-and-biases approach in several ways. First, they argued that the approach was not ecologically valid, in the sense that it did not follow Brunswik's (1947) principle of representative design, in which the stimuli presented are representative of those in the environment. In the current literature, this criticism mixes up two issues. One is that heuristics may be much more useful than they seem in HB experiments, which are often based on unusual situations such as those faced by juries or physicians. Indeed, even the use of simple representative designs indicates that heuristics can come very close to normative models (Read & Grushka-Cockayne 2011). HB proponents might well agree but also argue that the unusual situations are important even if they are not typical.

The second issue is to claim that heuristics were at some point adaptive, for example, some earlier point in the biological evolution of humans, so, therefore, they are not biases. This argument seems to me to be without merit. For one thing, it is difficult to study environments faced by early humans, so we are free to make up stories about them. For another, HB proponents are more interested in what we can do now to improve the achievement of our goals. To say that our problems were once not problems at all is little help in solving them.

A second criticism is that HB proponents rely too heavily on coherence and do not pay enough attention to correspondence. In weather forecasting, for example, forecaster A could say, frequently, that the probability of rain is .6 and the probability of no rain is .6, while forecaster B says that the probability of rain is .1 and the probability of no rain is .9, on the same days. If it rains on half of the days, forecaster A is clearly better in terms of correspondence, but she is incoherent, while forecaster B is coherent. Isn't correspondence what matters?

I have two answers to this second criticism. First, HB researchers study decisions much more than fast-and-frugal researchers, and correspondence is of little use in the study of decisions. Second, in the field of judgment, the criticism might be true but a great deal of research in the JDM field uses correspondence criteria in the study of judgment (if only correspondence with "the right answer"), so the criticism is based on selection of those studies to which it applies.

A third criticism is that simple heuristics sometimes lead to judgments that conform to normative models better than attempts to follow the normative models directly. This is simply true. A dramatic demonstration of the power of simple heuristics is the recognition heuristic. Goldstein and Gigerenzer (2002, 2011) asked subjects which of two cities was larger. The subjects were German, and they did better with American cities than German ones, because they did not recognize many of the American ones. When an American recognized city was paired with an unrecognized city, subjects chose the one they recognized. This was a very good cue, and German subjects used it to their advantage.

Many studies have examined the use of heuristics in judgment tasks in which subjects predict some quantity, or quantitative comparison, on the basis of several other cues. An example is predicting college grades from high-school grades and several test scores. Many studies use artificial examples in which subjects must learn from experience how the target is related to the predictors (cues). In most real-life and laboratory studies, the relation between cues and target was well captured by a linear additive model. The best way to predict the target was to use a weighted average, that is, multiply each cue by its optimal weight and add up the results (adding a constant term because of differences in scales). But the subjects did not know the optimal weights. And, in fact, any attempt to determine the optimal weights from a limited sample of cases would get them wrong: some would be too high and others too low. Thus, the strategy of *trying* to use a weighted average would suffer from this problem of “overfitting” the data. Under certain conditions, accuracy of prediction can be improved by reducing the variation in weights toward equal weighting. (Equal weighting means that each predictor is first standardized, and then the simple sum, plus a constant, is used to predict the target.) It is as if we need to compromise between what the (limited) data tell us and what we might think before we had the data, which is that we would weight the predictors equally. Sometimes this effect is so extreme that simple equal weighting does better than an attempt to find the optimal weights.

Equal weighting is one of a class of simple strategies that might be used in tasks of this sort. Others are simple heuristics, such as the recognition heuristic, or “Take the best,” for comparing two items. To apply take-the-best, go through the attributes in order from best to worst. If one item is better on an attribute, choose it. If the two are equal, proceed to the next attribute. If you run out of attributes, guess. Another heuristic is simply to use the best single cue. Because of the general problem of overfitting when attempting to use a weighted-average strategy, most heuristics like these can, under certain conditions, do better than an explicit attempt to compute a weighted average (Dana & Dawes 2004; Hogarth & Karelaia 2007; Davis-Stober, Dana, & Budescu 2010).

Putting this another way, the use of simple heuristics can be normatively optimal by the criterion of correspondence, under specific conditions. It helps, of course, if people know when each heuristic will be most useful. (This is discussed in the articles just cited, and their citations.) This conclusion goes beyond the argument that sometimes it is worthwhile to sacrifice a little accuracy for the benefit of being “fast and frugal.” In some cases, accuracy need not be sacrificed at all, because it improves with the use of simple heuristics.

This happens because the attempt to apply what seems to be the obvious normative model (weighted average) directly is prone to error. Yet, normatively, we should take the possibility of error into account in our judgments. (In general, we could apply EU theory to do this, looking at the probability of various errors and the reduction in utility—the disutility—that they cause.) Hence, it may be argued that the apparent normative model is not normative at all, because it does not take error into account.

In this regard, the argument is similar to an argument made in moral philosophy by (for example) J. S. Mill (1859). If you have the power to suppress free speech, there

are certainly situations in which it would be best (in terms of overall consequences) to do so. The trouble is that we make errors in trying to recognize those situations, and the disutility of these errors is very high when we mistakenly act to suppress speech. Moreover, the difficulty of recognizing such errors is indicated by the poor track record of people who have made this distinction in the past. Thus, Mill argued, we do best by following a simple rule of never suppressing speech, regardless of how the situation appears.<sup>3</sup> And we can do this knowingly, without deceiving ourselves into believing that we are doing anything other than seeking the best overall consequences.

### 3 OTHER HEURISTICS

---

A great many heuristics have been discovered, other than those just described. Baron (2008) discusses the following, among others:

- Missing information: do not choose an option if you lack information about its consequences (even if you cannot get the information and the option is otherwise good).
- Congruence: choose tests of hypotheses that will give yes answer if your favored hypothesis is true.
- Information: seek information (even if you don't know what to do with it when you get it).
- Equality: divide benefits or burdens equally.
- Do no harm: avoid harming others, even when doing so can prevent greater harm.
- Status quo: stick with the status quo (similar to "Choose the default").
- Proportionality: evaluate interventions in terms of the proportion of the problem that they correct.
- Reciprocity: reciprocate harms and benefits as closely as possible.
- Waste: do not waste.
- Variety: seek variety.

Each of these is associated with a bias, but other heuristics are prescriptive, such as "Consider alternative hypotheses" or "Ask what you will do with the information you get."

<sup>3</sup> Although Mill did make an argument like this against suppressing free speech, another, more current, examples is the decision to commit an act of terrorism. Surely some terrorist actions have been for the best, but practically every terrorist believed that his act is in that category, while the overwhelming majority have been wrong. Thus, any potential terrorist should correct his own belief by taking base rates into account.



Close examination of various tasks can reveal many more heuristics specific to the tasks of interest. For example, Hauser (2011) reviews heuristics used by consumers in choosing products. One heuristic is to pick a small set of the available options before considering more details. (This one has obvious timesaving advantages and less obvious advantages in reducing error rates.) Another is to infer quality from an advertising campaign (which is rational but fallible: suppliers with “money to burn” can afford such campaigns). Another is to use popularity as a cue, or recognition (which is of course one of the goals of advertising).

Similarly, Sunstein (2005) describes a number of “moral heuristics,” which affect judgments about public policies.

## 4 VIEWS OF HEURISTICS

---

Theoretical use of the idea of heuristics grew from accretion of examples, not from any clearly stated theory. Perhaps as a result of this history, no single theory seems to account for all the heuristics that have been described, although a few different theoretical approaches can describe significant groups of heuristics.

### 4.1 Development and Overgeneralization

It is potentially useful to study the development of biases in children and the role of heuristics, but few studies have done this, and those that have attempted it have not reached enlightening general conclusions (e.g., Baron et al. 1993). The problem, as noted by Stanovich, West, and Toplak (2011), is that this kind of research is beset with conceptual and methodological difficulties. For example, measurement of the frequency of biases in a population may depend heavily on other aspects of difficulty of the task, such as how familiar the situation is. Young children are less familiar with almost everything, so they may appear to be more biased for spurious reasons.

However, an interesting possibility is that some heuristics arise from learning, followed by more learning to reduce them, so they show an inverted-U pattern with age. When children learn to talk, they often learn general rules and overapply them. Very young children do not use the plural form of nouns. Older children learn that the plural is formed by adding an “s” sound, and some children apply this everywhere, saying “mouses” instead of “mice.” As they get older still, they learn the exceptions, the limits of the simple rules. Arkes and Ayton (1999) suggested that the same thing happens with the sunk-cost effect, the tendency to stick to a losing strategy when the future consequences of giving up are better. (This bias seems to be explained in terms of a “Do not waste” heuristic, as if the waste could be redeemed by wasting even more time or money.) The youngest children show no bias. Older children show the bias most strongly. Adults show it much less because they have learned that the do-not-waste rule

should be limited to cases in which the waste has not yet occurred. The same process can be demonstrated in learning by adults, such as the overgeneralization of the law of large numbers by students of statistics (Ganzach & Krantz 1991, pp. 189–90).<sup>4</sup>

The possibility of this kind of result suggests that some heuristics can be viewed as over-generalizations of useful rules. For example, we decide on punishment on the basis of the magnitude of the harm done, we stick to the status quo (“A bird in the hand is worth two in the bush”), and we judge harmful acts to be worse than equally harmful omissions. These heuristic rules are discovered by individuals and passed on from one person to another because they work in certain situations, perhaps in most situations. Their success can be understood when their results are best at achieving the thinker’s goals. For example, it is usually best to punish harmful acts more than we punish harmful omissions, because harmful acts are usually more intentional and their outcomes are easier to foresee, so they are more easily deterred. But people use these heuristics without fully understanding their relationship to their purposes. As a result, people often apply them in cases in which they do not achieve their usual purposes as well as some alternative might. Thus, people sometimes judge acts to be worse than omission even when intention and foreseeability are held constant (chapter 3 by Baron in this volume).

In some cases, people can reflect on the rules they use and realize that the rules have exceptions (Baron & Leshner 2000). In other cases, especially those that involve socially acquired rules such as those concerning morality, people seem to become committed to the rules they use, even though reflection about their purposes might reveal limitations, if people could imagine what purposes might justify the rules.

The term “overgeneralization” is somewhat misleading. Overgeneralization of one rule goes hand in hand with *undergeneralization* of whatever rule should be used in its place. However, the term is still useful because it brings to mind other examples of inappropriate transfer of a rule. Overgeneralization errors were taken by Wertheimer (1959) to be a sign of misunderstanding, of “blind” learning or transfer. Wertheimer showed that such overgeneralization can apply to rules of inference learned in school as well as “naive” rules. For example, he found that students who had learned the formula for the area of a parallelogram (base times height) would apply the same formula to other figures that could not be made into rectangles in the same way that a parallelogram can be (by cutting off a triangle on one side and moving it to the other). The area rule worked for the parallelogram because it could be shown in this way to serve the purpose or goal of finding area. The derivation of the rule involves the subgoals of making the parallelogram into a rectangle and conserving area. When the rule was applied

<sup>4</sup> Students were asked, “A person is standing in the middle of the road after drinking all evening. He is so drunk that his steps are completely random. He is as likely to make a step to the right as he is to make a step to the left.” Is he more likely to be closer to the point of origin after 100 steps or after 1,000? Students who had studied statistics were *less* likely to give the correct answer (100), presumably because they applied a heuristic based on the law of large numbers, which they had just learned.

elsewhere, it did not serve the purpose of finding area, and the subgoal of making the figure into a rectangle could not be achieved.

The possibility that some heuristics are overgeneralization errors suggests a general kind of reflection that might reduce biases that they cause. In these cases, the overgeneralization of a rule is incorrect because it no longer serves the purpose of the rule when applied correctly. If people “searched for goals,” that is, for purpose, and thought about the goals that rules helped to achieve, they might be able to learn limitations, such as the overuse of the do-not-waste rule. Some rules may turn out to be purposeless.

## 4.2 Attribute Substitution and the Affect Heuristic

Another way of account for a great many heuristics is that they involve attribute substitution: “When confronted with a difficult question people often answer an easier one instead” (Kahneman & Frederick 2002). When asked about probability of category, people answer in terms of similarity (thus ignoring the base rates of the categories). When asked about an appropriate jail term for a crime, people answer by thinking about how outrageous the crime is.

This account applies to some of the same heuristics that can be explained in terms of overgeneralization. For instance, in the proportionality effect in judgments of priority for risk reduction. The relevant measure is (for example) the number of deaths prevented, but people sometimes find it easier to think about the proportion of deaths prevented. This is obviously an attribute substitution, but it could also be described as the application of an incorrect rule, a rule that usually works well (judge in terms of proportion). Although the attribute-substitution account is more obviously applicable here, that is not the case for other heuristics, such as “Do no harm,” which seems more like a rule, although it could also be described in terms of attribute substitution (substituting “harms from action” for “total harms”).

One attribute that is often substituted, or used when it should be irrelevant is “affect,” hence the *affect heuristic* (Slovic et al. 2002). This term has a specific technical meaning, which is an immediate evaluative judgment: good-bad, happy-sad, and so on. It does not seem to mean “emotion” in a technical sense of that term. The difference is that emotion is a mental state that itself has hedonic value, while affect is a judgment, which may or may not be accompanied by emotion. The word “murder” may evoke an affective judgment by association (murder-bad), but need not evoke any emotion by itself.

## 4.3 Isolation Effects

Yet another way of accounting for some heuristics is that they involve attention to what is present in the immediate field of consciousness, neglecting what is outside but easily accessible with some mental search. This is called isolation effect, or the focusing effect. The idea came from the theory of mental models in reasoning (as explained

by Legrenzi, Girotto, & Johnson-Laird 1993): people reason from mental models, and when possible they use a single, simple, model that represents just the information they are given. Other factors are ignored or underused. Other statements of essentially the same idea are those of Read, Loewenstein, and Rabin (1999) and Idson and coauthors (2004).

An example of this is Thaler's (1985) "mental accounts." Thaler found that many people treat the source of funds as relevant to their use, although money is fungible. Thus people who are normally frugal and even risk averse would spend lottery proceeds on luxury items or binge purchases. In doing so, they viewed their windfall gains in isolation, and failed to integrate their newfound wealth with all their liabilities and assets. (See chapter 23 for another example, from the work of McCaffery and Baron (2006) on tax.)

Isolation effects may be subsumed under availability effects. They could also be described in terms of attribute substitution, where the isolated attribute is substituted for all those attributes that should be considered. And they could be described as over-generalized rules. But they are specific to the situation where the attribute that is used is the most salient, if only because it is part of the question that people are asked.

## 4.4 Two-Systems (Dual Process) Theory

The idea that there are two systems of cognitive processes has received considerable attention since the publication of Kahneman's (2011) *Thinking, Fast and Slow*. Kahneman proposes that a fast, automatic, associative process is initially evoked by the need for a judgment or decision, System 1. It is often followed by a slower, more reflective, rule-governed, and controlled process, System 2, which may serve to correct errors made by system-1. Many harmful heuristics work at System 1. Although Kahneman's book has increased attention to the dual-process idea, it has been around in psychology since the nineteenth century and has taken many forms, which differ in their statements about what is included in each system or type of process.

Recent expositions include those of Epstein (1994), Evans (2008), Frederick (2005), Greene (2009), Slovic (1996), and Slovic (2007), among others. The main dimension of differences among the various versions is the amount that is included in each system. At the most extreme, System 1 is described as fast, automatic, associative, unconscious, not limited in what it can do at once, affected by emotion, driven by heuristics, and more. System 2 is the opposite. It seems doubtful that such a clear division exists. At the opposite extreme, System 1 is simply automatic and System 2 is controlled.

Extensive evidence supports the simpler distinction. In the Stroop (1935/1992) task, for example, subjects are asked to name the color in which words are printed. They have great difficulty with this when the words are themselves the names of colors. Presumably, reading the word is the automatic or natural response, and people must suppress it to do the task. In Frederick's (2005) cognitive-reflection test, subjects are given problems like, "If it takes 5 machines 5 minutes to make 5 widgets, how long

would it take 100 machines to make 100 widgets?” The natural response of 100 must be suppressed. It is unclear whether the suppression is done in advance, by making a plan not to consider the natural response, or only after the initial, incorrect, response is almost made. In the latter case, then we would have to conclude that the System 1 response is always faster, but we would not need to assume this in advance suppression can be accomplished (although it may still be true).

Although this distinction is clear, what is not clear is whether heuristics generally operate automatically in the same way as giving 100 as the answer to the widget problem. Clearly Polya’s answer would be no. Heuristics are often intentionally deployed. As Polya pointed out, and Gigerenzer said in a different way, some heuristics are learned methods that improve performance. These heuristics may be intentionally deployed. And the heuristics that causes biases may also be intentionally deployed.

Many biases that are presumably caused by heuristics seem to exist even when subjects are required to reflect on them (e.g., Frisch 1993). In some cases, it seems that heuristics, or principles that are labeled as such, take the form of principles that are endorsed after reflection. This may occur in the moral domain (Sunstein 2005), and also in decisions and judgments based on probability. Despite serious normative arguments against the relevance of “ambiguity” or “uncertainty about probability” to decision-making (Baron & Frisch 1994), many scholars have defended the principle of avoiding options when the “probability of some outcome is unknown,” perhaps because the intuition that it is relevant is very powerful (Slovic & Tversky 1974).

Another example is the neglect of base rates usually attributed to the representativeness heuristic. Such neglect is found in experimental subjects, but it is also sometimes defended by judges in their opinions (Koehler 2002).

## 4.5 Debiasing Heuristics

People can learn to avoid harmful heuristics. This learning can result from ordinary education, or from special training directed at the heuristics themselves. For example, Fong, Krantz, and Nisbett (1986, experiment 4) found that statistical training transfers to solving everyday problems involving statistical reasoning, in a telephone interview in which subjects were asked about regression to the mean, such as why the winner of the Rookie of the Year award in baseball usually does not do as well in his second year as in his first. A nonstatistical response might be “because he’s resting on his laurels; he’s not trying as hard in his second year.” A statistical response (based on the principle of regression to the mean) would be “A player’s performance varies from year to year. Sometimes you have good years and sometimes you have bad years. The player who won the Rookie of the Year award had an exceptional year. He’ll probably do better than average his second year, but not as well as he did when he was a rookie.” Students gave more good statistical answers of this sort at the end of the course than at the beginning. Therefore, these students did transfer what they had learned to cases where it is relevant. Failure to regress to the mean—the assumption that the deviation from

average would be just as great in the second year—is possibly caused by the representativeness heuristic.

Nisbett and coauthors (1987) carried out other studies in which they examined the effects of various kinds of university instruction on statistical, logical, and methodological reasoning. They found that logical reasoning did not improve from the beginning to the end of two introductory college courses in logic (one emphasizing formal logic, the other emphasizing informal fallacies). It did not improve as a result of two years of graduate school in chemistry, either. Logical reasoning did improve, however, as a result of two years of graduate school in law, medicine, or psychology. The researchers suggest that these three fields emphasize logical reasoning to some extent. Statistical and methodological reasoning showed a somewhat different pattern. The largest improvement by far occurred among psychology students, probably because training in such things as the use of control groups is an important part of graduate work in psychology. Methodological reasoning also improved with medical training, which places some emphasis on the conduct and interpretation of medical research, but there was essentially no improvement from training in either law or chemistry.

Jury instructions might be designed to reduce biases. An example is suggested by McCaffery, Kahneman, and Spitzer (1995) in the context of tort damages.

For some heuristics based on attribute substitution or overgeneralization, a general method of debiasing might be helpful, although it has not been tried. That is to encourage people to ask themselves about the purposes (goals) of the methods that they use (Baron 1993). This intervention might be most effective when users of these heuristics do not think they are doing anything nonnormative; they either think that they are making correct inferences or they do not think about the validity of what they are doing at all. In the contrasting case, the user knows that the heuristic is a shortcut and uses it to save time or effort. Here, intervention might be less useful without also convincing people that they needed to spend more time.

## 5 MY-SIDE BIAS AND ACTIVELY OPEN-MINDED THINKING

---

Another possible general approach to reducing biases involves a large category of biases, which are called “my-side bias” (Perkins, Bushey, & Faraday 1986). These biases favor whatever possibilities are already strong (options, beliefs, or goals). They thus lead to irrational persistence in the face of counterevidence or counterarguments. Beliefs defended in this way are often but not always what the person wants to be true, so my-side bias is not the same as wishful thinking. Depressives are subject to these biases when they interpret evidence as favoring the hypothesis that bad outcomes are their fault. (See Baron 2008, for a more complete review.)

My-side bias exists in both search and inference. In search, it is a bias to look for reasons that favor the stronger possibility. For example, political liberals (conservatives) tend to read articles written by liberals (conservatives). Then, having selected the evidence by its content, they ignore the fact that they did this when they revise their beliefs, so they change their beliefs as if the evidence were selected randomly. They do not discount the evidence for the fact that it was selected in a biased way. (If they did, they would probably not spend so much time searching for it, for its expected benefits would be small.) The same sort of selective search is involved when the evidence comes from our own memory. We tend to think of evidence that supports possibilities that are already strong or beliefs that we desire to be true, and we need to put aside our knowledge that we have done this, in order to take the result seriously.

Several experiments indicate that this sort of bias underlies both motivated and unmotivated errors. In one commonly used experimental paradigm, subjects are instructed to make some judgment under various conditions: without special instructions, with instructions to think of reasons why their initial judgment might be correct, with instructions to think of why it might be incorrect, or with instructions to think of reasons on both sides. (Not all experiments use all four conditions.) The typical finding is that the latter two conditions improve performance over the first two. For example, Koriat, Lichtenstein, and Fischhoff (1980) found that thinking of reasons on the “other side” reduced inappropriate extreme confidence in the answers to objective questions. Results like this suggest that one source of many of these errors is a bias toward initial or desired conclusions. Telling people to think of evidence for these conclusions has no effect, because they do it anyway. Telling them to think of counterevidence helps, whether or not we tell them to think of supporting evidence too.

Perkins, Bushey, and Faraday (1986) has provided additional evidence for the existence of such my-side bias. When subjects were asked to list arguments relevant to some question of policy, such as whether their state should enact a law requiring return of glass bottles for recycling, they listed far more arguments on their own side than on the other side of the issue. When they were pushed to list *extra* arguments, they listed far more such arguments on the *other* side. So their initial bias was the result of lack of effort, not lack of knowledge.

The way to counteract my-side bias is a prescriptive heuristic, or set of heuristics, called actively open-minded thinking (AOT: Baron 2008). This involves actively seeking evidence and arguments against strong possibilities (beliefs, options, or goals), including other possibilities that might be considered.

Some evidence suggests that AOT can be taught. Beginning with Selz (1935), several studies have shown that something resembling actively open-minded thinking can be trained and that it transfers to new situations (Baron, Badgio, & Gaskins 1986; Perkins, Bushey, & Faraday 1986). Such training does not need to be outside of regular courses. It can often be done by changing the design of traditional course, often in ways that even facilitate the learning of the course content (Baron 1993).



## 6 NAIVE THEORIES

---

Unfairly neglected in recent literature is another approach, pursued mostly in the study of child development, which holds that people have naive theories. (See Wellman & Gelman 1992, for a review.) Children think that the sun goes around the earth, and that the earth is flat. Naive theories are also found in adults. Some people think that a thermostat works like an accelerator, so that, if you turn it way up, the room will heat up faster (Kempton 1986). This approach is somewhat like the idea of biases, as compared to normative models. But the “true theories” to which judgments are compared are not seen so much as definitions of optimality that are good for all time but rather as just the most defensible current thinking. (Really the distinction is not that sharp.) People undoubtedly have naive theories of economics, of how the law works, and of their roles as citizens (Baron 2012).

The study of naive theories can be contrasted with the study of biases relative to normative models. In the latter case, we can say that people are biased, nonnormative, or perhaps irrational. In the former case, we can say that they are incorrect. Just as we must accept normative models in order to say that people are biased, we must accept current understanding of science (or law, or whatever) to say that people are incorrect, so any conclusions are conditional on such assumptions. Despite this limitation, study of naive theories could open up new approaches to the improvement of human thinking, understanding, and choice.

One area in which naive theories are both abundant and consequential is consumer finance. Recent debates over regulation of financial markets have cited a number of behavioral anomalies that are arguably caused by naive theories, especially in the high-stakes context of retirement planning. For example, Benartzi and Thaler (2001) documented the phenomenon of “naive diversification,” showing that most people think that portfolio diversification is essentially an investment of  $1/N$  into each of  $N$  funds offered—they believe that spreading out investments is good, irrespective of the attributes of the options at hand. Similarly, many investors rely on an explicit theory that favors funds with strong past performance and ignores fund fees, which are deemed too small to matter (Wilcox 2003), though in fact most retail investors are better off ignoring past performance and choosing the lowest-fee funds available.

## 7 CONCLUSION

---

It has long been assumed by many scholars, other than some logicians and economists, that people are not entirely rational in any strict sense of the term. As a result of research over mostly the last 50 years, we now have a much better idea of just where people deviate from simple theories of rational thinking or decision-making.

These are optimistic results. If we can trace the problems of individuals and the world to specific kinds of poor thinking, then we open a path to fixing these problems, a path that would not be available if the problems arose entirely from other causes. There are several things we can do.

One is to try to teach people to do better. For the kinds of biases I have discussed, such instruction is probably best done in tertiary education (college, in the United States). Primary and secondary schools could do it, but the prospects for changing the curriculum based on mere evidence, in my experience, are dim.

Another is to rely on decision architecture (Thaler & Sunstein 2008) to present decisions to people in ways that will help them make better decisions for themselves and others. As Thaler and Sunstein point out, full use of this approach will require changes in the law and its application in the courts and by government and nongovernmental agencies. This approach is paternalistic. Thaler and Sunstein argue that it is “libertarian paternalism” because people are still free to choose. But, as pointed out by Camerer and coauthors (2003), many proposals of the sort that Thaler and Sunstein recommend actually impose some cost on people, if only the cost of making a special effort to take some option other than the default. Some proposals, such as waiting periods to avoid impulsive decisions (about marriage or divorce, for example) impose a noticeable cost on everyone. Camerer and coauthors argue that these costs can be justified by the benefits they yield in terms of harm prevention. The extreme case, of course, is outright paternalism resulting from legal prohibition or requirements, which can also be justified in economic/utilitarian terms (Zamir 1998).

A final approach is to reduce the scope for biases to operate by providing people with better information. Again, Sunstein and Thaler recommend this, and sometimes it must be legally required (as in warning labels). Just to take one example, a great many biases have been discovered in the domain of risk perception, and these seem to affect risk regulation (Breyer 1993), given the fact that citizens, legislators, and regulators are all subject to the same biases. But many of these biases have greater scope to operate when people are ignorant of the facts about the costs and benefits of various possible regulations. So publish the facts as best we know them. (This has still not been done. There is no authoritative website to which one can turn for a summary of this sort of information, except in limited domains such as healthcare.) With the numbers in hand, people would find it easier to think in ways that are approximately normative. As the late John McCarthy said, “He who refuses to do arithmetic is doomed to talk nonsense.”

## REFERENCES

- Arkes, H. R., & Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin* 125, 591–600.
- Baron, J. (1985). *Rationality and intelligence*. New York: Cambridge University Press.
- Baron, J. (1990). Thinking about consequences. *Journal of Moral Education* 19, 77–87.

- Baron, J. (1993). Why teach thinking?—An essay. (Target article with commentary.) *Applied Psychology: An International Review* 42, 191–237.
- Baron, J. (2004). Normative models of judgment and decision making. In D. J. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making*, pp. 19–36. London: Blackwell.
- Baron, J. (2008). *Thinking and deciding* (4th edition). New York: Cambridge University Press.
- Baron, J. (2012). The “culture of honor” in citizens’ concepts of their duty as voters. *Rationality and Society* 24, 37–72.
- Baron, J., Badgio, P., & Gaskins, I. W. (1986). Cognitive style and its improvement: A normative approach. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*, vol. 3, pp. 173–220. Hillsdale, NJ: Erlbaum.
- Baron, J., & Frisch, D. (1994). Ambiguous probabilities and the paradoxes of expected utility. In G. Wright & P. Ayton (Eds.), *Subjective probability*, pp. 273–94. Chichester, Sussex: Wiley.
- Baron, J., Granato, L., Spranca, M., & Teubal, E. (1993). Decision making biases in children and early adolescents: Exploratory studies. *Merrill Palmer Quarterly* 39, 23–47.
- Baron, J., & Leshner, S. (2000). How serious are expressions of protected values? *Journal of Experimental Psychology: Applied* 6, 183–94.
- Bell, D. E., Raiffa, H., & Tversky, A. (Eds.) (1988). *Decision making: Descriptive, normative, and prescriptive interactions*. New York: Cambridge University Press.
- Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in retirement saving plans. *American Economic Review* 91, 79–98.
- Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review* 115, 463–501.
- Breyer, S. (1993). *Breaking the vicious circle: Toward effective risk regulation*. Cambridge, MA: Harvard University Press.
- Bröder, A., & Newell, B. R. (2008). Challenging some common beliefs: Empirical work within the adaptive toolbox metaphor. *Judgment and Decision Making* 3, 205–14.
- Brunswik, E. (1947). *Systematic and representative design of psychological experiments. With results in physical and social perception*. Berkeley: University of California Press.
- Camerer, C. F., Issacharoff, S., Loewenstein, G., O’Donoghue, T., & Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for “asymmetric paternalism.” *University of Pennsylvania Law Review* 151, 1211–54.
- Dana, J., & Dawes, R. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics* 29, 317–31.
- Davidson, D., Suppes, P., & Siegel, S. (1957). *Decision making: An experimental approach*. Stanford, CA: Stanford University Press.
- Davis-Stober, C. P., Dana, J., & Budescu, D. V. (2010). Why recognition is rational: Optimality results on single-variable decision rules. *Judgment and Decision Making* 5, 216–29.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist* 49, 709–24.
- Ericsson, K. A., & Ward, P. (2007). Capturing the naturally occurring superior performance of experts in the laboratory: Toward a science of expert and exceptional performance. *Current Directions in Psychological Science* 16, 346–50.
- Evans, J. St. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology* 59, 255–78.
- Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology* 18, 253–92.

- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives* 19, 25–42.
- Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature* 40, 351–401.
- Freeling, A. N. S. (1984). A philosophical basis for decision aiding. *Theory and Decision* 16, 179–206.
- Frisch, D. (1993). Reasons for framing effects. *Organizational Behavior and Human Decision Processes* 54, 399–429.
- Ganzach, Y., & Krantz, D. H. (1991). The psychology of moderate prediction: I. Experience with multiple determination. *Organizational Behavior and Human Decision Processes* 47, 177–204.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.
- Glöckner, A. (2009). Investigating intuitive and deliberate processes statistically: The multiple-measure maximum likelihood strategy classification method. *Judgment and Decision Making* 4, 186–99.
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review* 109, 75–90.
- Goldstein, D. G., & Gigerenzer, G. (2011). The beauty of simple models: Themes in recognition heuristic research. *Judgment and Decision Making* 6, 392–95.
- Greene, J. D. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology* 45, 581–84.
- Hauser, J. (2011). A marketing science perspective on recognition-based heuristics (and the fast-and-frugal paradigm). *Judgment and Decision Making* 6, 396–408.
- Hogarth, R., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review* 114, 733–58.
- Idson, L. C., Chugh, D., Bereby-Meyer, Y., Moran, S., Grosskopf, B., & Bazerman, M. (2004). Overcoming focusing failures in competitive environments. *Journal of Behavioral Decision Making* 17, 159–72.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Strauss and Giroux.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment*, pp. 49–81. New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3, 430–54.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review* 80, 237–51.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47, 263–91.
- Kempton, W. (1986). Two theories of home heat control. *Cognition* 10, 75–90.
- Koehler, J. J. (2002). When do courts think base rate statistics are relevant? *Jurimetrics* 42, 373–402.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory* 6, 107–18.
- Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making*, pp. 316–37. London: Blackwell.

- Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (1993). Focusing in reasoning and decision making. *Cognition* 49, 36–66.
- McCaffery, E. J., & Baron J. (2006). Isolation effects and the neglect of indirect effects of fiscal policies. *Journal of Behavioral Decision Making* 19, 1–14.
- McCaffery, E.J., Kahneman, D., & Spitzer, M. L. (1995). Framing the jury: Cognitive perspectives on pain and suffering awards. *Virginia Law Review* 81, 1341–1420.
- Mill, J. S. (1859). *On liberty*. London: J. W. Parker & Son.
- Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The Drift Diffusion Model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judgment and Decision Making* 5, 437–49.
- Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science* 238, 625–31.
- Perkins, D., Bushey, B., & Faraday, M. (1986). Learning to reason. Final report, Grant No. NIE-G-83-0028, Project No. 030717. Harvard Graduate School of Education.
- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin* 68, 29–46.
- Polya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton, NJ: Princeton University Press.
- Read, D. & Grushka-Cockayne, Y. (2011). The similarity heuristic. *Journal of Behavioral Decision Making* 24, 23–46.
- Read, D., Loewenstein, G. L., & Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty* 19, 171–97.
- Selz, O. (1935). Versuche zur Hebung des Intelligenzniveaus: Ein Beitrag zur Theorie der Intelligenz und ihrer erziehlichen Beeinflussung. *Zeitschrift für Psychologie* 134, 236–301.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin* 119, 3–22.
- Slovic, P. (2007). “If I look at the mass I will never act”: Psychic numbing and genocide. *Judgment and Decision Making* 2, 79–95.
- Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*, pp. 397–420. New York: Cambridge University Press.
- Slovic, P., & Tversky, A. (1974). Who accepts Savage’s axioms? *Behavioral Science* 14, 368–73.
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review* 31, 103–18.
- Stroop, J. R. (1935/1992). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General* 121, 15–23.
- Sunstein, C. R. (2005). Moral heuristics (with commentary). *Behavioral and Brain Sciences* 28, 531–73.
- Tetlock, P. E. (2005), *Expert political judgment: How good is it? How can we know?* Princeton, NJ: Princeton University Press.
- Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science* 4, 199–214.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.
- Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*, pp. 153–60. Cambridge: Cambridge University Press

- Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology* 43, 337–75.
- Wertheimer, M. 1945 [1959]. *Productive thinking* (rev. ed.). New York: Harper & Row.
- Wilcox, R. T. (2003). Bargain hunting or star gazing? Investors' preferences for stock mutual funds. *Journal of Business* 76, 645–64.
- Zamir, E. (1998). The efficiency of paternalism. *Virginia Law Review* 84, 229–86.

## CHAPTER 2

---

# HUMAN PROSOCIAL MOTIVATION AND THE MAINTENANCE OF SOCIAL ORDER

---

SIMON GÄCHTER

## 1 INTRODUCTION

---

ECONOMIC approaches to understanding human behavior, including law abidance, have long assumed that people are self-regarding in the sense that they entertain cost-benefit calculations with the sole concern being own costs and benefits, irrespective of consequences for others. The last twenty years of research in behavioral economics have profoundly challenged this assumption (Gintis et al. 2005) with important consequences for our understanding of lawful behavior and social order in general. The question I will discuss in this chapter is how prosocial motivations help understand social order.

I will discuss evidence from the last two decades of behavioral economics research that sheds light on human prosocial motivations. I will focus my attention on people's behavior in social settings where the welfare of other people is affected. Of course, there are important behavioral aspects of law abidance from an asocial, individual decision-making perspective. These concern the roles of probability perception (e.g., the perceived probability of being caught for a criminal act) and heuristics and biases in general (Kahneman and Tversky 2000). I do not deal with these issues here but refer the interested reader to Sunstein (2000), and to chapter 1 by Baron in this volume. My focus is on prosocial motivation, not cognition.

The basic conceptual framework I will use to study social order goes back to at least Hobbes and consists in thinking of social order as a cooperation problem: If the law is widely disregarded, we end up in a world where life is "nasty, brutish, and short." Contributing to social order (obeying the law) is of collective interest, but individuals have an incentive to disregard the law if this promises to be more advantageous



than abiding by the law. Of course, the law also has an important coordinating function that makes obeying the law also in people's self-interest—think of traffic laws, for instance. In this chapter, however, I will not discuss coordination problems, but focus on cooperation.

One may argue that modern societies rest on constitutions and legal enforcement mechanisms with checks and balances that sharply limit individual decisions to flout the law. However, as examples from failed states vividly demonstrate, legal enforcement cannot work if large groups in society disregard the law because they think respecting the law is not to their advantage. Thus, a functioning law enforcement system is itself an example of successful cooperation.

The problem of cooperation can most easily be understood in the following simple example. Suppose two farmers reach an agreement to respect each other's property, but there is no third party to enforce this agreement. Now the two farmers have to think whether to stick to their promise or not. Suppose they both abide by their agreement and therefore have an incentive to cultivate their land, which gives them a comfortable living. But one farmer might be tempted to renege on his promise and steal the harvest of the other farmer, who trusted that he would be safe and therefore invested in a good crop. The stealing farmer enjoys a harvest for which he did not work, and the victim is robbed of his proceeds. If farmers are not gullible (or learn from experience), they might anticipate this outcome and not invest much in cultivating the land, which leaves both in a miserable situation but still better off than losing the entire harvest after a season of hard work. This, of course, is the famous prisoner's dilemma: mutually sticking to the agreement is in the common interest but not in each individual's interest.

The prisoners' dilemma as a metaphor for cooperation has been the focus of decades of research (Rapoport and Chamah 1965; Axelrod 1984; Van Lange et al. 2014). One important insight has been that cooperation (i.e., honoring the agreement) might be maintained in my example if these farmers are likely to play the game in the foreseeable future with the "shadow of the future" (Axelrod 1984) looming strongly enough. The mutual threat to renege on the agreement if the other farmer reneges might be strong enough to cause both to honor the agreement. If the agreement is successful, we have an example of self-enforcement. Such self-enforcement is much harder to achieve if the players are not settled farmers but mobile bands of hunter-gatherers because under the latter conditions there is no common shadow of the future but only a short-term cooperation problem, which favors defection.

Modern social life differs, of course, from this simple example: cooperation problems need to be solved for large groups, where decisions take place both in stable settings and random interactions. But large groups, even if they are stable, are fundamentally different from two-person prisoner's dilemmas: theory suggests (Boyd and Richerson 1988) and experimental evidence confirms (Grujić et al. 2012) that stable cooperation is possible in the two-person prisoner's dilemma but is hard to achieve in large groups because no effective punishment targeted at noncompliant group members exists. Thus, for understanding large-scale cooperation the prisoner's dilemma is not fully suitable, and recent research has therefore shifted to the public goods game as



a tool to study multilateral cooperation. This game will be the major tool I will use in this chapter.

One important insight of many experiments using the public goods game is that cooperation is inherently unstable and tends to unravel to the worst outcome, predicted by self-interest. Doesn't this prove that people are selfish in the end? My answer will be a qualified no. Some people are indeed likely to be selfish. Many people, however, will *behave* selfishly under some conditions, but are not *motivated* by selfishness. As I will show, the distinction between motivation and behavior is important and they ought not to be conflated. People can be nonselfishly motivated and end up behaving selfishly, but the converse also exists: selfish people behaving prosocially.

The main tool to investigate my questions is economic experiments, with decision-dependent monetary stakes. A full description of the experimental methodology is beyond the scope of this chapter. The interested reader should consult Falk and Heckman (2009) and chapter 5 by Engel in this volume.

The plan of this chapter is as follows. Section 2 will lay the foundation of my analysis of determinants of social order by offering an overview of the most important findings suggesting that the *homo economicus* assumption used for decades in economics and other behavioral sciences is not justified. Many people are more aptly described as *homo reciprocans*, that is, nonselfish "strong reciprocators" (Gintis 2000), and I will present the most important evidence supporting the existence of strong reciprocators. A strong reciprocator is prepared to sacrifice resources to be kind to those who are being kind ("strong positive reciprocity") and to punish those who are being unkind ("strong negative reciprocity"). The essential feature of strong reciprocity is a willingness to reward fair and punish unfair behavior even if this is costly and provides neither present nor future material rewards for the reciprocator (Fehr, Fischbacher, and Gächter 2002). However, as I will show, all experiments that find evidence for strong reciprocity also find the existence of mostly self-regarding people.

The rest of this chapter will then discuss how *homo economicus* and *homo reciprocans* deal with social order. I will argue that social order is sustained, to some extent, by internalized norms of proper conduct even in the absence of any formal enforcement. Social order is also, and very strongly so, influenced by the behavior of other people because *homo reciprocans* is more likely to contribute to the common good if others do the same. I will also show that punishment or other incentives are necessary to sustain social order.

A first pillar of social order, and probably the weakest one, is personal ethics, or internalized norms of cooperation, enforced by feelings of guilt: Cooperation can be supported to the extent that people think cooperating is the morally right thing to do and feel guilty if breaking the social contract. Section 3 investigates the role of internalized norms of proper conduct to sustain cooperation. In section 4 I will show that social order is bound to be fragile if not backed up by incentives. This holds despite the fact that most people are not fundamentally self-regarding and, as section 3 will show, express moral apprehension at free riding. An important insight is that some people are selfish and that *homo reciprocans*, while not being selfish, sometimes tends

to be selfishly biased. Section 5 discusses evidence that (the threat of) punishment is crucial to maintain social order. *Homo reciprocans* has a decisive role to play because *homo reciprocans* is prepared to pay a cost to punish those who jeopardize social order. Rewards and a desire for a good reputation can also help.

Section 6 will present some cross-societal evidence and show that punishment is also shaped by how well the rule of law works in a given society. Section 7 will present a short discussion and outlook for future research.

Before I proceed I should clarify what this chapter does and does not provide. Research in the behavioral sciences searches for basic behavioral principles that underlie all social dilemmas however diverse they are in reality. My approach therefore is not applied science (although I will point to some interesting applied findings) but basic science that should provide general behavioral principles that can inform more applied research. The behavioral research I report here is complementary to approaches studying the role of social norms in law and its enforcement (e.g., Ellickson 1991; Posner 2000; Kahan 2003).

## 2 BASIC SOCIAL MOTIVATIONS: *HOMO ECONOMICUS* AND *HOMO RECIPROCANS*

---

*Homo economicus* has long been the most important characterization of human nature in the behavioral sciences and in particular in economics. David Hume famously remarked that “Political writers have established it as a maxim, that, in contriving any system of government, and fixing the several checks and controls of the constitution, every man ought to be supposed a knave, and to have no other end, in all his actions, than private interest” (Hume 1987 [1777], Essay VI, p. 42). George Stigler, a Nobel laureate in economic sciences, was convinced: “Let me predict the outcome of the systematic and comprehensive testing of behavior in situations where self-interest and ethical values with wide verbal allegiance are in conflict. Much of the time, most of the time in fact, the self-interest theory. . . will win” (Stigler 1981, p. 175).

There are several justifications for the selfishness assumption. *Homo economicus* is neutral to other people, that is, he is neither envious or malicious and also not altruistic. Thus, he might be considered the average person on whom social analysis should be based (Kirchgässner 2008). Furthermore, in a theoretical context, the *homo economicus* assumption often allows for exact predictions, which can be confronted with appropriate data that might refute it. Moreover, it is often of independent interest to understand what would happen if everyone were self-regarding. A clear picture of the consequences of selfishness serves therefore as an important benchmark for understanding nonselfish behavior.

The assumption of self-regard also has considerable merit in the absence of empirical means to assess the structure of people’s social preferences. Yet the experimental

methodology allows us to observe people's social preferences under controlled circumstances. Advances in neuroscience (Glimcher et al. 2009), anthropology (Henrich et al. 2004), behavioral economics (Gintis et al. 2005), evolutionary theory (Bowles and Gintis 2011) and social psychology (Van Lange et al. 2014) shed further light on human nature. Thus, given the availability of appropriate tools to measure deviations from selfishness, there is no need to rely further on the selfishness assumption. Its empirical relevance can now be measured.

In the following I will present evidence that supports the widespread existence of *homo reciprocans*. The classic games used to study people's social preferences are the dictator game, the ultimatum game, the trust and the gift-exchange game, and the public goods game with and without punishment. All experiments I will discuss are conducted according to the standards of experimental economics (see Friedman and Sunder 1994 for a textbook account) and have been replicated many times, including in representative samples, under high stakes, and in relevant field conditions. Moreover, all experiments are designed to carefully control for self-regarding incentives, such that self-interest theory makes a unique prediction that can be compared with the behavioral outcome. If behavior differs from the self-interest prediction we have evidence for nonselfish behavior.

The *dictator game* (Forsythe et al. 1994) is the most basic decision situation in which social preferences can be studied. The dictator game is a two-player game where participants are assigned at random to be either a "dictator" or a passive recipient. The dictator has to decide how much of a given amount of money allocated to him or her to share with a recipient who has to accept the offer. The experimental setting ensures that a self-interested rational dictator has an incentive not to share. Passing money along to the recipient under these conditions is evidence for altruism, or other-regarding preferences in general.

The results of many carefully controlled dictator games do not support self-interest predictions on average. In a meta-analysis Engel (2011) finds that across 616 treatments involving the dictator game, the average sharing rate is 28.3%, and across all studies about 36% of individuals do not share at all. Thus, many people are willing to share a windfall gain, but (depending on the treatment) a sizable minority is not.

How about sharing principles if recipients can reject the offer? The seminal game to study this situation is the *ultimatum game* (Güth, Schmittberger, and Schwarze 1982). In the ultimatum game the proposer makes an offer of how to share a given pie and, in contrast to the dictator game, the recipient can now accept or reject the offer. In case of acceptance, the offered division is implemented; in case the recipient rejects, both get nothing. If the recipient is motivated solely by monetary payoffs, he or she will accept every offer. Therefore, the proposer will only offer the smallest money unit.

The results across a wide range of subject pools around the world reject this prediction (Oosterbeek, Sloof, and van de Kuilen 2004). On average, proposers offer 30% to 40% of the available amount. The median and the mode are at 40% and 50%, respectively. Few offers are less than 10%, or more than 50%. Offers below 20% or less will likely be rejected, while equal splits are almost always accepted.

The offers made in the ultimatum game appear inconsistent with the *homo economicus* model of human nature. However, it is important to observe that all types of proposers, self- and other-regarding ones, have an incentive to offer nonminimal (fair) shares, if some recipients are inclined to reject low offers. Thus the mere fact that we observe high offers is not inconsistent with the *homo economicus* model. The inconsistency arises for the recipient who foregoes earnings by rejecting a positive offer—*homo economicus* would never do that. Cross-societal variation notwithstanding, when it comes to rejections, there is abundant support for the existence of strong negative reciprocity, and no support for the *homo economicus* prediction in almost any of the many societies studied (Oosterbeek et al. 2004; Henrich et al. 2005; Henrich et al. 2006).

The next game, the *gift-exchange game* (developed by Fehr, Kirchsteiger, and Riedl 1993), showcases strong positive reciprocity where *homo reciprocans* nonstrategically rewards a kind act by being kind as well. A simple version of the gift-exchange game works as follows. There are two roles, employers and employees. In each round, an employer and employee are paired up at random. The employer makes a wage offer to his or her paired employee, who can accept or reject the offer. Acceptance concludes an employment contract. The employee then chooses effort and the round ends. “Effort” means choosing a number with the consequence that the higher the chosen number, the higher is the employer’s profit and the higher is the employee’s effort cost. The earnings of employers increase in effort and decrease in wages paid. For the employee the opposite holds. Parameters are such that maximal effort maximizes surplus.

The setup ensures that there are no strategic reasons for gift exchange. A *homo economicus* employee will choose the minimum effort irrespective of the wage because effort is costly. *Homo reciprocans*, however, will respond reciprocally: high wages are rewarded with high effort and low wages are matched with low effort.

The results of numerous experiments support the *homo reciprocans* prediction over the *homo economicus* one, on average, because wage and effort are highly significantly correlated. This is unambiguous evidence for strong positive reciprocity, found in numerous gift-exchange experiments (see Charness and Kuhn 2011 for an overview). However, the results also reveal substantial heterogeneity. Irrespective of the wage paid by the firm there is always a fraction of workers who choose minimal effort—as in the dictator game, *homo economicus* exists but is in the minority.

A game related to the gift-exchange game that also allows for the observation of strong positive reciprocity is the *trust game* (Berg, Dickhaut, and McCabe 1995). The trust game is a two-player game where participants are anonymously and at random allocated to their roles as trustor and trustee. The trustor (and in some experiments also the trustee) has an endowment and has to decide how much of this endowment to transfer to the trustee. Any amount the trustor transfers the experimenter increases by a factor of three (in some studies by a factor of two or four). The trustee then decides how much of the increased amount to transfer back to the trustor. *Homo economicus* in the role of recipient will not return anything irrespective of the amount received (and rational trustors would foresee this and transfer nothing).

Numerous studies with student and wide-ranging nonstudent subject pools have been conducted with the trust game. Johnson and Mislin (2011) found in a meta-analysis of 162 replications in 35 countries that trustors on average send 50% of their endowment and trustees return 37% of the amount available for return. Regression analyses show clear support for strong positive reciprocity: Trustees return highly significantly more the more they have received from the trustor.

In sum, the gift-exchange game and the trust game provide substantial evidence for the existence of strong positive reciprocity. Rejections in the ultimatum game are an example of strong negative reciprocity. But these are all two-player games. I have argued in the introduction that to understand human cooperation one needs to move beyond dyadic interactions. The following game, the public goods game, is a vehicle to study strong positive reciprocity in the context of a simultaneous multi-lateral game.

In a typical *linear public goods game*, four people form a group. All group members are endowed with 20 tokens. Each member has to decide independently how many tokens (between 0 and 20) to contribute to a common project (the public good). The contributions of the whole group are summed up. The experimenter then multiplies the sum of contributions by a factor larger than one but less than four (a frequently used factor is 1.6) and distributes the resulting amount equally among the four group members irrespective of how much an individual has contributed. Thus, an individual benefits from the contributions of other group members, even if he or she has contributed nothing to the public good. A rational and self-regarding individual has an incentive to keep all tokens, because the personal benefit per token from the public good is less than 1, whereas it is 1 if he or she keeps the token. By contrast, the group as a whole is best off if everybody contributes all 20 tokens.

A large number of studies show that people contribute to the public good (see Chaudhuri 2011 for an overview), but, as I will describe in more detail in the next section, contributions decrease over time in experiments that allow for repetition of the base game. In this section I focus on one-shot games because my goal is to demonstrate the existence of strong positive reciprocity, and this requires controlling for any self-regarding incentives. One-shot games provide the starkest environment to study cooperation motivated by strong reciprocity because there are no strategic reasons to make any positive contribution. Thus, *homo economicus* will take a free ride in this game. Again, the results do not confirm this prediction. Many people make a positive contribution, although a significant fraction contributes nothing (e.g., Dufwenberg, Gächter, and Hennig-Schmidt 2011).

The fact that people make positive contributions does not yet constitute evidence for strong positive reciprocity. In a game where group members make their contribution decisions simultaneously people cannot respond to what they have observed others to do; people can only react to the *beliefs* they hold about other group members' contributions. Thus, some experiments ask the participants what they estimate the other group members will contribute (e.g., Dufwenberg, Gächter, and Hennig-Schmidt 2011). The results are consistent with strong positive reciprocity: on average, reported beliefs and

own contributions are highly significantly positively correlated. While this holds for a majority of people, some contribute nothing despite the fact they believe others will contribute a lot. Again, *homo economicus* and *homo reciprocans* coexist.

The final game I discuss is the *public goods game with punishment* (Fehr and Gächter 2000; Fehr and Gächter 2002). It provides another example for the existence of strong negative reciprocity. In this game, the group members first contribute to the public good. Then group members learn how much all have contributed and are given the opportunity to spend money to reduce the income of each of the other group members individually. One money unit spent on punishing a group member reduces this group member's earnings from the first stage by three money units. *Homo economicus* will of course not spend any money to punish others, but *homo reciprocans* might be willing to punish the free riders in the group.

The results show that many people are prepared to punish free riders. In fact, in the experiments of Fehr and Gächter (2002) more than 80% of people punished at least once. Fehr and Gächter repeated their experiment six times, but each time with entirely new group members and in a way that excluded any further interactions with any previous group members. Punishment showed a reciprocal pattern in each of the six one-shot repetitions: more free riding was met with more punishment. Gächter and Herrmann (2009) and Cubitt, Drouvelis, and Gächter (2011) found the same result in strict one-shot experiments. Such punishment has been called "altruistic" because it is individually costly and benefits others only; it is evidence of strong negative reciprocity. These experiments and a related large literature (surveyed in Gächter and Herrmann 2009; Balliet, Mulder, and Van Lange 2011; and Chaudhuri 2011) show that many people are strong negative reciprocators with punitive sentiments for wrongdoing.

What are possible psychological (proximate) mechanisms that produce strong reciprocity? At the most fundamental level, it is the evolved human capacity of empathy that only psychopaths lack (Baron-Cohen 2011). Relevant for my specific question, research has identified three important mechanisms: inequality aversion (Fehr and Schmidt 1999; Bolton and Ockenfels 2000); efficiency seeking (Charness and Rabin 2002); and a desire to reward or punish intentions behind actions (also called reciprocity; Rabin 1993; Dufwenberg and Kirchsteiger 2004) or a combination of inequality aversion and rewarding and punishing intentions (Falk and Fischbacher 2006). A detailed description is beyond the scope of this chapter, but I will provide the ideas. Thorough discussions can be found in the cited articles and in Fehr and Schmidt (2006); see Wilkinson and Klaes (2012) for a textbook account. Bowles and Gintis (2011) provide evolutionary (ultimate) explanations of strong reciprocity.

*Inequality aversion.* Inequality aversion, in particular the version by Fehr and Schmidt (1999), is probably the most widely used theory to explain the reviewed behavior in experimental games. The theory assumes that people care about their own material payoff positively and negatively about inequality in comparison with another person both in case the inequality is advantageous (the focal individual has more than the comparison individual) and if it is disadvantageous (the focal individual has less



than the comparison individual). It is assumed that disadvantageous inequality is worse than advantageous inequality.

The theory of inequality aversion can explain why people reject unfair offers in ultimatum games (an example of strong negative reciprocity): while there is a positive utility from the material benefit of a (small) offer, there is also disutility from inequality. It is therefore possible that the disutility outweighs the utility from the offered amount, and therefore total utility is negative and the person rejects. A second example of how disadvantageous inequality aversion can explain strong negative reciprocity is punishment of free riders in a public goods game: a free rider who does not contribute anything will earn more than all others who have contributed. This will leave the contributors behind in payoff comparisons and they will experience disadvantageous inequality aversion. A contributor who punishes a free rider may reduce the gap in earnings and therefore inequality by punishing. Aversion to advantageous inequality can also explain why people behave in a positive reciprocal way when making contributions to a public good. If a group member believes others will contribute he or she might feel advantageous inequality aversion if not contributing. To alleviate this feeling, she contributes. An inequality averse person will also not contribute more than others because this way she would fall behind in terms of payoffs.

There are, however, a couple of important phenomena that the theory of inequality aversion does not address: many people are motivated by efficiency seeking and are therefore willing to help others even if that increases inequality (a strictly inequality averse person would not do that), and people care not only about outcomes as assumed in theories of inequality aversion, but also about the intentions behind actions. I will deal with these two problems in turn.

*Efficiency seeking.* Inequality aversion implies that people will always take actions, if available, that reduce inequality. But experiments have shown that many people are also willing to help other people if that increases efficiency despite also increasing inequality (Charness and Rabin 2002; Engelmann and Strobel 2004). Thus, a social concern for efficiency most likely is an important motivation for some, and it might also explain why people make contributions to public goods.

*Intentions matter.* A second problem with the theory of inequality aversion is that it is purely outcome-oriented, that is, the *intentions* behind other people's actions do not matter. However, there are many cases where intentionality is important. For example, receiving an unfair offer (involving a disadvantageous unequal distribution) if a fair offer (an equal distribution) is available might not be perceived the same as if the only available other offer is also unfair (Falk, Fehr, and Fischbacher 2003). Theories of reciprocity (e.g., Dufwenberg and Kirchsteiger 2004) model intentions by assuming that people are motivated by rewarding kindness with kindness and meanness by meanness. Making a fair offer when an unfair offer is available (and better for the proposer) is an example of a kind act; offering an unfair distribution when a fair one would have been available is an example of unkind behavior. Another example is contributions to a public good: if a group member believes others will contribute a lot, then

he or she might perceive this as a kind act and reward the kindness by contributing as well; by the same token, a low expected contribution might be perceived as unkind and therefore be matched with a low contribution as well (Dufwenberg, Gächter, and Hennig-Schmidt 2011). Falk and Fischbacher (2006) combine inequality aversion and intentions, and show that intentions might lead to more punishment of unfair offers and free riding than inequality aversion alone.

In sum, strong positive and negative reciprocity are probably to a large extent motivated by psychological mechanisms of inequality aversion and a desire to base rewards and punishments on the intentions behind an action; in some important cases concerns for social efficiency also matter. Existing research clearly suggests an important role for these mechanisms (Falk, Fehr, and Fischbacher 2005). For my purposes, however, it suffices to work with strong positive and negative reciprocity as motivational shortcuts.

In the following I will turn to the central question of this chapter, the determinants of social order. In the next three sections I will show how the basic inclinations of strong positive and negative reciprocity determine (the breakdown of) social order.

### 3 THE DETERMINANTS OF SOCIAL ORDER

#### I: INTERNALIZED NORMS

---

One important determinant of people's prosocial behavior is most likely internalized norms of what people consider the morally right thing to do. For example, people donate anonymously to charities (Eckel and Grossman 1996), they vote for reasons of civic duty, despite their vote being extremely unlikely to be pivotal (Riker and Ordeshook 1968); they respect the law (Cooter 2000) even if incentives that back up the obligations are weak (Galbiati and Vertova 2008). People pay their taxes despite low detection probabilities for evasion (Kirchler 2007), and people also care for the environment out of moral convictions (Brekke, Kipperberg, and Nyborg 2010). More generally, people value character virtues such as honesty and trustworthiness even if lying and cheating go entirely undetected (e.g., Gneezy 2005; López-Pérez and Spiegelman 2013; see Pruckner and Sausgruber 2013 for an interesting field study) and also act on perceived moral obligations (Schwartz 1977). As shown above, in experiments people make contributions to one-shot public goods without any extrinsic incentive to do so. One early piece of evidence that is consistent with intrinsic motivations is that people contribute for reasons of "warm glow" (Andreoni 1990).

In this section I discuss some evidence about normative considerations and related moral emotions in social dilemmas. I discuss studies that investigate people's moral judgments, the social emotions of anger and guilt, and people's desire to punish norm violators even if not personally affected ("third-party punishment").



I start with moral judgments of free riding. Is free riding morally blameworthy at all? Cubitt, Drouvelis, Gächter, and Kabalin (2011) report on a study that elicited people's moral judgments of free riding by using techniques from moral psychology to understand to what extent free riding is perceived to be a *moral* problem. The basic design of Cubitt and his colleagues' study is as follows. They presented their subjects—who took the roles of spectators—with scenarios of two people, A and B, who are both endowed with 20 money units and make contributions to a public good. B always free rides, that is, keeps all of his 20 money units for himself. The different scenarios vary the extent to which A makes contributions to the public good. Depending on the scenario, A contributes 0, 5, 10, 15, or 20 to the public good. People were asked, as a detached observer, how they morally judge B's behavior for each of A's possible contributions. The moral judgment scale ranged from  $-50$  (extremely bad) to  $+50$  (extremely good).

Figure 2.1A illustrates the result by showing the average moral evaluation of B's free riding (contribution of 0 to the public good) for each of A's possible contributions. The average moral evaluation is always below 0; that is, people think that B's free riding is morally blameworthy. Interestingly, the same act of free riding is considered morally worse on average the more A actually contributes.

Figure 2.1A shows the average moral evaluation, which hides some interesting heterogeneity. About 50% of people actually have a flat "moral judgment function"; that is, their moral evaluation of B's free riding does not depend on how much A contributes. A third of the people think B's free riding becomes morally worse the more A contributes.

If free riding is considered morally blameworthy, does it also trigger negative emotions? Evidence using noninvolved spectators who evaluate free riding behavior described in various scenarios, suggest so (Fehr and Gächter 2002). And if cooperation is morally commendable, does free riding trigger feelings of guilt? Anger and guilt are

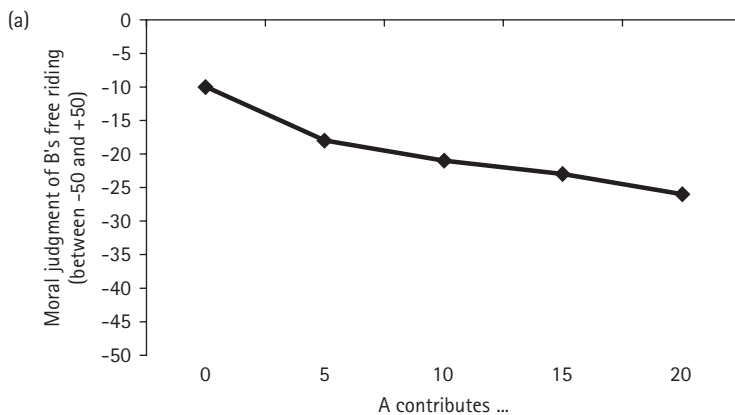


FIGURE 2.1A Moral Judgment of Free Riding

Data source: Cubitt et al. (2011); own illustration.

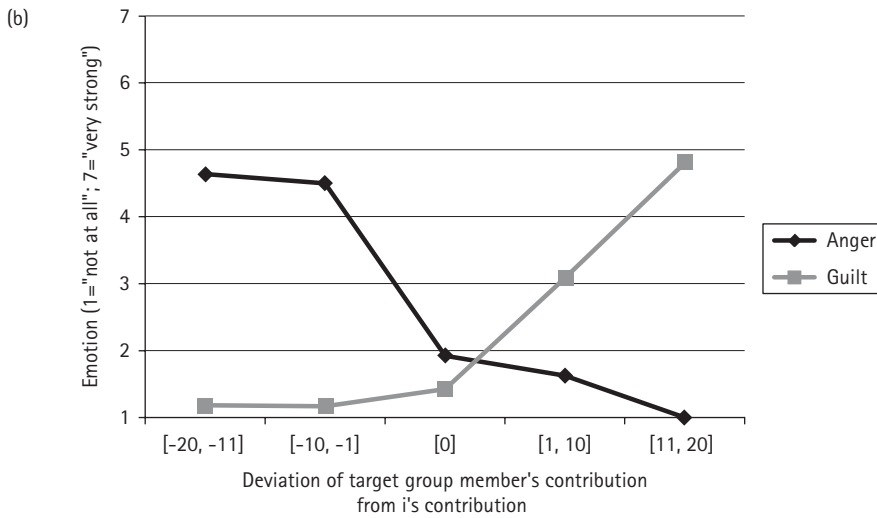


FIGURE 2.1B Moral Emotions: Anger and Guilt

Data source: Cubitt, Drouvelis, and Gächter (2011); own illustration.

expected to be particularly relevant in a context of social cooperation because they can be seen as prototypical morally linked emotions (e.g., Haidt 2003).

Cubitt, Drouvelis, and Gächter (2011) elicited emotions after players made their contributions in a one-shot public good to see to what extent free riding triggers anger by the cooperating individual and guilt by the free rider. Figure 2.1B shows that the average levels of anger and guilt seem to be mirror images of one another with the exception that a high level of free riding (where the target individual contributes between 11 and 20 tokens less than the focus individual) triggers the same anger as a lower level of free riding (the target individual contributes between 1 and 10 tokens less than the focus individual).

The moral or social emotions anger and guilt are interesting because they trigger two potential enforcement mechanisms—external and internal punishment. Angry individuals might be willing to punish free riders and therefore provide the free riders with an extrinsic self-regarding incentive to avoid punishment by contributing (discussed in more detail in section 6). Guilt is a negative emotion that can serve as “internal punishment” and therefore provide an intrinsic reason to contribute to the public good to avoid feeling guilty. Dufwenberg, Gächter, and Hennig-Schmidt (2011) presented evidence that such “guilt aversion” can explain contributions to public goods.

Further evidence for the importance of normative considerations comes from third-party punishment games (Fehr and Fischbacher 2004), where a potential punisher is not an affected party, but an independent third party (this feature thus resembles law enforcement in reality). In their experiment, two players, A and B, play a prisoner’s dilemma game with two options: Cooperate (C) or Defect (D). In terms of material payoffs, the best outcome for a player is DC, that is to defect when the other

player cooperates; the second-best outcome is CC (mutual cooperation); the third-best result is DD (mutual defection) and the worst outcome is CD (cooperating while the other player defects). This incentive structure gives both players an incentive to defect and therefore to forgo the gains from mutual cooperation. Fehr and Fischbacher (2004) add to this framework a third party who, at her own cost, can punish both players A and B after having seen their decisions. Since the third party is not affected by A's and B's decisions, third-party punishment is a reflection of normative considerations. The results show that third parties are much more likely to punish a defector if the other player cooperated (in 46% of cases) than if both defected (21% of cases); mutual cooperation is almost never punished.

The results on third-party punishment are consistent with the findings on moral judgments (figure 2.1A): free riding is considered particularly blameworthy if the other party cooperated. The third-party experiments uncovered that people have a willingness to pay for their normative convictions. Neuroscientific evidence (Buckholtz and Marois 2012) as well as cross-cultural findings (Henrich et al. 2006) suggest that third-party punishment is a phenomenon that is deeply ingrained in the human condition.

In summary, people think free riding is morally blameworthy and it also triggers the contributors' anger and even third-party punishment. People who contribute less than others feel guilty. Thus, to the extent that people have feelings of warm glow, are bound by moral norms, want to avoid making other group members angry even if (third-party) punishment is not possible, and would feel guilty if contributing less than others, prosocial cooperation is expected.

## 4 THE DETERMINANTS OF SOCIAL ORDER II: THE BEHAVIOR OF OTHER PEOPLE

---

I introduced the *one-shot* public goods game in section 2 as one tool to study the existence of strong positive reciprocity. The evidence suggests that people are willing to contribute to public goods even in one-shot settings. To investigate (the stability of) social order, however, requires repeated public goods games. Notice that the repeated public goods game is a stark setting in which to study social order: while one-shot settings allow observing people's principal willingness to cooperate for the sake of the collective benefit, a repeated setup allows answering the question whether this willingness can help producing a *stable* social order.

Are people able to provide a public good that has a collective benefit to all, if the collective benefit and the "shadow of the future" are the sole incentives? The fact that many people are guilt-averse, think free riding is immoral, and are also motivated by efficiency-seeking should help in pursuing collective welfare. However, the sobering result of many repeatedly played public goods experiments is that cooperation almost

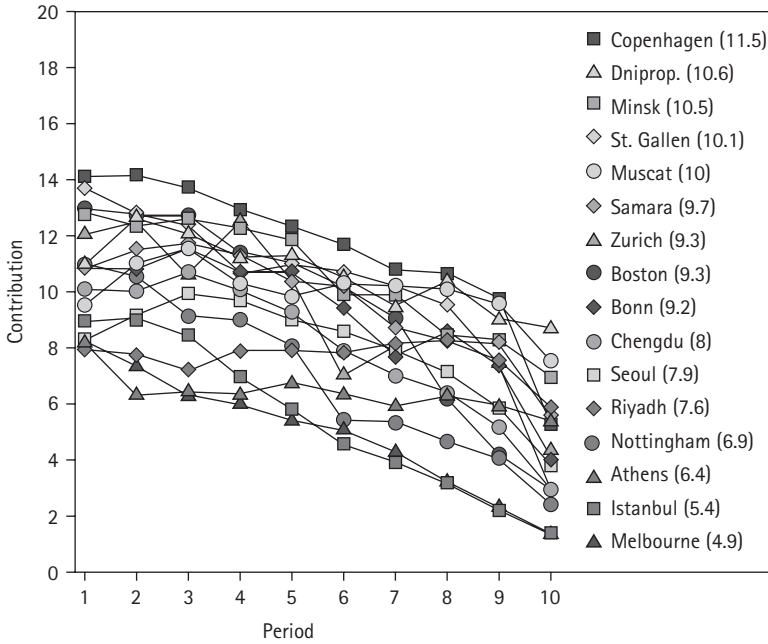


FIGURE 2.2 The Breakdown of Cooperation Is Ubiquitous: Evidence from Fifteen Countries

Source: Herrmann, Thöni, and Gächter. 2008.

Figure 2.2 shows the average contribution (out of the endowment of 20 money units) the subjects contributed in each round. The numbers in parentheses are average contributions over all rounds.

invariably breaks down in repeated interactions. This result has been shown in numerous experiments around the world (Herrmann, Thöni, and Gächter 2008; Chaudhuri 2011) and is illustrated in figure 2.2. In all subject pools people contribute substantial amounts initially but over time contributions dwindle to low levels almost everywhere.

One may argue that the experiments reported in figure 2.2 are too short to properly reflect conditions relevant for social order. Unfortunately, existing experimental evidence suggests that the time horizon does not matter much. For example, in Gächter, Renner, and Sefton (2008) participants played for ten or fifty periods (and participants knew this). Cooperation was low under both time horizons (less than 40% on average) but not different between time horizons. Rand and coauthors (2009) and Grujić and coauthors (2012) report very similar results (Grujić et al. even for 100 periods, and, as in Rand et al. 2009 with participants being unaware of the exact number of rounds). Thus, the conclusion is inevitable and seems to vindicate Hobbes: in and of itself, that is, without external enforcement, social order is fragile and the time horizon as such is of no avail.

Recall from section 2 that one-shot public goods experiments have found a positive correlation between beliefs about the contributions of other group members and the individual’s own contributions, which is consistent with strong positive reciprocity. However, this positive correlation is not a particularly compelling measure of strong

positive reciprocity. To see why, suppose, for whatever reason, Alice is very pessimistic about the contributions of others and thinks they will not contribute much or may contribute nothing at all. Suppose further Alice would be willing to contribute provided others also contribute—Alice is a “conditional cooperator.” Alice behaves as a free rider due to her pessimism, not because her basic attitude to cooperation is free riding. Now compare Alice to Bill and assume that Bill is a free rider who will never contribute even if others contribute a lot. Thus, there is a problem: Alice and Bill both free ride, so their *behavior* is observationally equivalent, but their *motivation* is different. Bill is motivated to be a free rider, whereas Alice is a conditional cooperator who happens to be pessimistic. Thus, separating behavior from motivation is important (see Lewinsohn-Zamir 1998 for a related argument in a law and public policy context).

Fischbacher, Gächter, and Fehr (2001) introduce a design that allows separating behavior from motivation. In their experiment, participants are asked in an incentive-compatible way to make conditional contributions for all possible average contributions of the other group members (a so-called strategy method). Given the details of the incentive structure, people motivated by free riding will contribute nothing for all levels of possible average contributions of other group members. Conditional cooperators, by contrast, will increase their contribution in line with the average contribution of others. Thus, in this design, rather than just observing one contribution and one belief, we can observe a complete contribution schedule for all possible average contributions of others. Free riders and conditional cooperators are therefore clearly distinguishable, even if they both contribute nothing if others contribute nothing. Fischbacher, Gächter, and Fehr (2001) find that about 50% of their participants are conditional cooperators, 30% are free riders, and the rest follow some other patterns. The average person clearly is a conditional cooperator.

The Fischbacher, Gächter, and Fehr (2001) experiment has been replicated many times in many countries and including representative subject pools (Thöni, Tyran, and Wengström 2012). Figure 2.3 illustrates the average conditional contribution from subjects in ten different countries around the world by showing the average contribution that subjects make as a function of all possible average contribution levels of other group members (expressed in percentages of the maximal possible contribution, which differs across studies).

A couple of important insights can be taken away from figure 2.3. First, although there is some variation, patterns are very similar across subject pools: low contributions by other group members are met with low own contributions, and own contributions increase in those of group members. This is true in all ten subject pools illustrated here. Second, while contributions increase in the contributions of others, own contributions tend to remain below the diagonal, which implies that even conditional cooperators on average want to free ride to some extent on the contributions of other group members. Figure 2.3 depicts average conditional cooperation and it therefore hides heterogeneity. However, conditional cooperators are the majority and free riders a minority in all

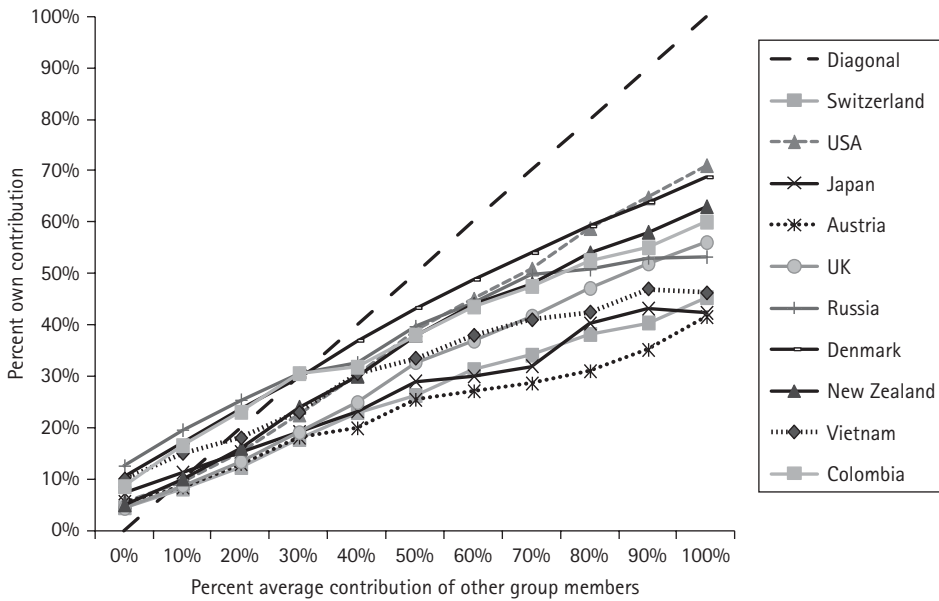


FIGURE 2.3 The Average Person Is a Conditional Cooperator: Evidence from Ten Countries

Data source: Chaudhuri and Paichayontvijit (2006) (New Zealand); Kocher et al. (2008) (Austria, Japan, USA); Herrmann and Thöni (2009) (Russia); Fischbacher, Gächter, and Quercia (2012) (UK, Switzerland); Thöni, Tyran, and Wengström (2012) (Denmark); Martinsson, Pham-Khanh, and Villegas-Palacio (2013) (Vietnam, Colombia). Own illustration.

subject pools studied. The assumption that *homo economicus* describes average behavior is thus not supported by the experimental findings.

Before I move on to discuss how the observation of figure 2.3 can explain the fragility of social order, it is worth discussing three more observations about conditional cooperation: several psychological mechanisms predict conditional cooperation, which makes it a highly likely pattern; conditional cooperation is externally valid; and conditional cooperation predicts contributions in experimental public goods games.

*Several psychological mechanisms support conditional cooperation.* Conditional cooperation is a likely pattern of behavior because various psychological mechanisms predict it. I already mentioned two proximate mechanisms of strong reciprocity in section 2—*inequality aversion* and a desire to match like with like (*reciprocity*). Numerous experiments suggest the existence of inequality aversion and reciprocity and I have already sketched the argument that these motivations can explain conditional cooperation. Conditional cooperation is also supported by cooperative *social value orientations*, where people take into account the welfare of others (Balliet, Parks, and Joireman 2009; Van Lange et al. 2014). A further channel to support conditional cooperation is *guilt aversion*, introduced in section 3. If Alice thinks others expect her to contribute, she might feel guilty if she doesn't and avoid feeling guilty she actually makes a contribution to the public good; if she expects others not to contribute, she will also not feel guilty by not contributing herself.

Moreover, conformism, a deep-rooted human tendency to copy other people's behavior, also supports conditional cooperation. A desire to conform will lead a conformist to contribute if he or she thinks that is what other people will do; of course conformists will also free ride if that is what the majority does. This argument has found some experimental support (Carpenter 2004).

*Conditional cooperation has external validity.* Conditional cooperation is not only observed under laboratory conditions but also in naturally occurring environments. For example, field experiments demonstrate donations to public goods consistent with conditional cooperation (e.g., Frey and Meier 2004; Shang and Croson 2009). Rustagi, Engel, and Kosfeld (2010) ran experiments with forest management groups in Ethiopia. They employ a measure similar to that used in the experiments summarized in figure 2.3 and show that groups with a high share of conditional cooperators are more successful in forest management (an important public good in Ethiopia) than groups with a higher share of free riders. A final example is tax morality, which displays the behavioral logic of conditional cooperation; that is, people are more likely to be honest in their tax declaration if they think most other people are as well (Frey and Torgler 2007; Traxler 2010).

*Conditional cooperation predicts contributions.* Conditional cooperation is not only a phenomenon with high external validity; it is also internally valid in the sense that the elicited cooperation preferences predict actual play in new public goods games: people classified as conditional cooperators also behave as conditional cooperators in a new public goods game and free riders tend to contribute nothing as predicted for them (Fischbacher, Gächter, and Quercia 2012). Moreover, when attitudes to cooperation are elicited multiple times, most people fall into the same type categorization each time, that is, conditional cooperation and free riding are intrapersonally stable attitudes (Volk, Thöni, and Ruigrok 2012). This observation supports evidence that people's other- or self-regarding behavior is consistent across games (Yamagishi et al. 2013).

These observations are important for explaining why social order in and of itself, that is, without further incentives, is inherently fragile. As figure 2.3 shows, the average person is a conditional cooperator, but detailed analyses show that some people are free rider types who never contribute. Moreover, on average, even conditional cooperators are selfishly biased. Most conditional cooperators will make a positive initial contribution to the public good and then take the average contribution of the other group members as the new benchmark. The fact that most conditional contributors are also selfishly biased will induce them to contribute less than the average next time, and therefore cooperation will almost inevitably unravel and finally most people will contribute little or nothing to the public good. This prediction is consistent with the evidence (see figure 2.2; Fischbacher and Gächter 2010 for a rigorous analysis; and Chaudhuri 2011 for a survey of this literature).

This result of the unraveling of cooperation due to selfishly biased conditional cooperation teaches us two important lessons. First, due to the process of conditionally cooperative reactions on others' contribution, many people will eventually behave like a free rider (contribute little to the public good) despite the fact that they are not



motivated by selfishness. Second, cooperation is inherently fragile, and needs some support through other mechanisms to be sustainable.

One assumption I have been making so far is that people are sorted at random into groups, and all experiments I discussed did in fact implement random group assignment. However, in reality, people can sometimes choose the social group they want to be in. Thus, the question is, does sorting help? The answer is a qualified yes. If people manage to sort into groups with strongly reciprocal conditional cooperators, then such groups are indeed able to maintain high levels of cooperation and can prevent its breakdown (Gächter and Thöni 2005). This observation is consistent with conditional cooperation: if others cooperate, conditional cooperators will cooperate too. But successful sorting requires that the cooperative types are indeed sorted together and are able to prevent free riders from entering (Ehrhart and Keser 1999) and can credibly signal their type (for a discussion of signaling from a law point of view see Posner 2000). These are quite stringent conditions that may or may not be satisfied in real social groups.

In summary, conditional cooperation is an important human motivation for many and, as numerous experiments have shown, a highly relevant determinant of social order. Thus, although conditional cooperation allows for the *possibility* of self-sustaining cooperation, it is unlikely that conditional cooperators manage to maintain high levels of cooperation. This is due to the existence of a substantial fraction of free riders and to the fact that even conditional cooperators typically display some selfish bias.

## 5 THE DETERMINANTS OF SOCIAL ORDER III: PUNISHMENT AND OTHER INCENTIVES

---

One important lesson from the research reported in the previous section on why social order is fragile is that the only way a cooperator can avoid being “suckered” is to reduce his or her cooperation, thereby punishing everyone, even other cooperators. This raises the question whether targeted punishment (whereby group members can identify a free rider and punish him or her) actually can solve the free rider problem and prevent the breakdown of cooperation. Mancur Olson, in a seminal analysis of the free rider problem in collective action, argued that “only a *separate and “selective” incentive* will stimulate a rational individual. . . to act in a group-oriented way.” Olson further noted that selective incentives “can be either negative or positive, in that they can either coerce by punishing those who fail to bear an allocated share of the costs of the group action, or they can be positive inducements offered to those who act in the group interest” (Olson 1965, p. 51, emphasis in original).

But who should apply these selective incentives? One answer is that in modern societies the legal system does the punishment. However, the state with its law enforcement institutions is a novel phenomenon on an evolutionary time scale. For a large part of



human history, social order needed to be sustained without central institutions. And even in modern times, self-governance is often necessary in many important social dilemmas (Ostrom 1990).

One element of self-governance is informal sanctions as applied by other group members (Ostrom, Walker, and Gardner 1992; this is sometimes also called peer punishment). But the problem is that punishing is itself a public good: If Alice punishes a free rider who then subsequently contributes his or her share, Bill will benefit also, even if he has not punished (and thereby behaves as a “second-order free rider”). If Bill is a *homo economicus* he will certainly not punish if punishment is costly and has no personal benefit for him (which is likely in many situations), but if Alice is a *homo reciprocans* she might punish even if punishment is costly. The evidence on strong negative reciprocity, reported in section 2, as well as the seminal studies by Yamagishi (1986) and Ostrom, Walker, and Gardner (1992) suggest many people are indeed willing to punish free riders and the second-order public goods problem is actually less of an issue. As reasoned above, free riders who fear punishment might have a selfish incentive to cooperate, and higher rates of cooperation should also convince conditional cooperators to keep cooperating.

Fehr and Gächter (2000) developed an experimental design to study punishment and cooperation in a sequence of ten one-shot (random group members—“Strangers”) and fixed group (“Partners”; same group members) public goods game—settings that correspond to different real-life interactions. The experiment proceeded as follows. Subjects first made their contributions to the public good, and then they entered a second stage, where they were informed about the individual contributions of all group members. Subjects could assign up to ten punishment points to each individual group member. Punishment was costly for the punishing subject and each punishment point received reduced the punished subject’s earnings from the first stage by 10%.

The results support the *homo reciprocans* hypothesis that people are willing to punish free riders and that punishment increases cooperation. In both the Stranger and Partner conditions contributions increased over time—contrary to the *homo economicus* prediction. There is a substantial difference in cooperation rates between Partners and Strangers. Partners contributed about 85% of their maximal contribution and Strangers about 58%. By comparison, without punishment cooperation rates under Partners and Strangers were 38% and 19%, respectively. The fact that in the presence of punishment opportunities contributions even increased over time in a Strangers setting is particularly astonishing.

What explains the difference in cooperation between the Partners and Strangers condition? One likely channel is that at the cooperation stage within stable groups an interaction effect exists between the availability of punishment and strategic reciprocity (reciprocity that is also in the self-interest of a free rider due to the repeated nature of the interaction). Repeated interaction and punishment are complementary instruments to stimulate contributions. If only direct reciprocity is possible, cooperation collapses. If only punishment is possible but groups are formed randomly

and hence direct reciprocity is not feasible, cooperation is stabilized at intermediate levels.

A theoretically interesting benchmark case of the Stranger condition is a situation where the likelihood of future interaction is zero, that is, groups interact only once in the same constellation. This situation is interesting, because evolutionary theories of cooperation (see Rand and Nowak 2013 for a succinct summary), predict no cooperation in this case. Therefore, Fehr and Gächter (2002) set up a so-called perfect stranger design where in each of the six repetitions all groups are composed of completely new members, and participants are aware of this. The results show again that cooperation increases over time when punishment is available.

The experiments of Fehr and Gächter (2000) and Fehr and Gächter (2002) also had a setting where subjects first played a condition with punishment and were then told that in a new condition the possibility of punishment would be removed. Again, the results show that punishment leads to high and stable cooperation rates. But when punishment is removed, cooperation collapses almost immediately and dwindles to low levels. This suggests that a cooperative benchmark is not enough to support cooperation if not supported by the possibility of punishment.

While cooperation differs strongly between Partner, Stranger, and Perfect Stranger conditions, punishment patterns are qualitatively and even quantitatively similar across rounds: the more a group member deviates from the average contribution of his or her group members the higher is the punishment that he or she will receive. These observations are remarkable given that cooperation levels differ strongly between conditions. The fact that strong reciprocators punish even under Perfect Stranger conditions and that this punishment induces free riders to increase their contributions makes punishment altruistic: the punisher only bears the costs of punishment, and because under Perfect Strangers the punisher will not meet the punished group member again, the benefits of increased cooperation accrue solely to the future group members of the punished subject.

The experiments I have discussed so far force participants by way of experimental design into a condition where punishment is or is not available. What do people choose if they have a choice between being subjected to a condition where punishment is available and one where punishment is ruled out? Gürerk, Irlenbusch, and Rockenbach (2006) studied this question and got an interesting result. Initially, people opt for the no-punishment environment, but soon they experience the problems of free riding. This experience changes their preferences and after a few more rounds the majority prefers an environment with punishment.

The proximate mechanisms behind altruistic punishment give an indication why punishment is not a second-order public good in practice. Punishment seems to be an impulse triggered by negative emotions and not much by forward-looking considerations (e.g., Casari and Luini 2012).

By now, there has been a lot of experimental and theoretical work on punishment and its effectiveness to stimulate cooperation. This literature is too voluminous to discuss here and I refer the interested reader to relevant surveys (Sigmund 2007; Gächter

and Herrmann 2009; Balliet, Mulder, and Van Lange 2011). I concentrate on five issues that are most relevant for my present purpose: the role of the severity of punishment and costs of punishment for the success of cooperation; punishment as a mere threat; imperfect observation and errors; institutionalized punishment; and incentives provided by rewards and reputation.

*Severity and costs of punishment.* The monitoring frequency and the severity of inflicted punishment matter for the effectiveness of punishment to stabilize (or increase) cooperation (Egas and Riedl 2008; Nikiforakis and Normann 2008). The more severe punishment is for the punished subject per unit of received punishment the higher are contributions. Although punishment is to a large extent nonstrategic, it follows cost-benefit considerations in the sense that punishment is less likely used the more costly it is for the punisher (e.g., Anderson and Putterman 2006). The fact that the level of cooperation corresponds to the severity of punishment suggests that low contributors respond strategically to the expected harm of punishment. If severe punishment is expected, free riders are deterred and cooperate. That is, although not pro-socially motivated, expected strong negative reciprocity can induce a selfishly motivated person to *behave* like a cooperater. Experiments by Shinada and Yamagishi (2007) also confirm the argument that increased cooperation by free riders through punishment strengthens the resolve of conditional cooperators to cooperate.

*Punishment as a mere threat.* One important characteristic feature of punishment is that it might not be used very often if people anticipate punishment and therefore try to avoid it through appropriate action. This is how law enforcement works in many instances. In the case of contributions to a public good, punishment is not necessary if people contribute at high levels and punishment might therefore simply act as a deterrent. This argument requires that punishment be a credible threat, that is, punishment indeed occurs if contributions are too low. If punishment is credible, then in equilibrium it will not happen very often. The existence of strong reciprocators suggests that some people are indeed willing to punish free riders, so punishment should be credible. After having received punishment, free riders typically increase their contributions, so punishment has the desired behavioral effect. But can punishment also work as a mere threat?

To study the question whether punishment can also work as a mere threat, Gächter, Renner, and Sefton (2008) extended the experiment to fifty periods. This should give plenty of time to establish punishment as a credible threat, and later on as a mere threat with very little actual punishment necessary to sustain high and stable contributions.

Figure 2.4 depicts cooperation with and without punishment. In the latter condition, cooperation is modest and slowly dwindling to low levels. In the condition with punishment cooperation approaches very high levels quickly. Consistent with the threat effect, punishment frequency is relatively high in the early phase of the experiment but approaches very low levels (less than 10%) in the second half of the experiment. Thus, punishment can exert its power as a mere threat effect, yet the threat has to be there. If punishment is impossible, cooperation breaks down.

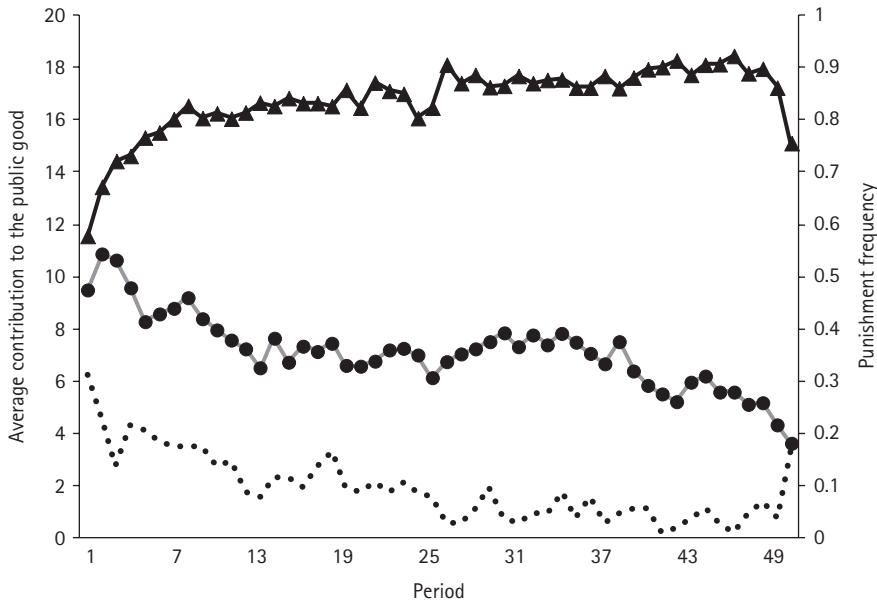


FIGURE 2.4 Punishment Can Stabilize Social Order through a Threat of Punishment Alone

Data source: Gächter, Renner, and Sefton (2008). Own illustration. Triangles indicate the punishment condition; circles indicate the no-punishment condition; and the dashed line indicates punishment frequency (measured on the right-hand vertical axis).

*Imperfect observation and errors.* All experiments I have discussed so far assume that all contributions are perfectly observable and no errors occur. This is quite unrealistic, and an important line of research investigates the consequences of imperfect observability and errors on punishment, cooperation, and overall efficiency of interactions. One way to model errors is to allow only for binary decisions: contribute or not (e.g., Ambrus and Greiner 2012). An error occurs if a contribution is actually registered as a noncontribution with a certain probability. If people apply the legal principle that punishment should only be used if the true act is known, little punishment of noncontributions should occur. However, a typical finding is that people punish too much and falsely hit a contributor too often with the consequence that punishment is less effective in stimulating cooperation than under perfect error-free observability of contributions (Bornstein and Weisel 2010; Grechenig, Nicklisch, and Thöni 2010). See Grechenig, Nicklisch, and Thöni (2010) for a discussion of the relevance of these findings from a legal science point of view.

*Institutionalized punishment.* The evidence I have discussed so far is all based on peer punishment. These experiments reveal two things: people get angry about free riders (see section 4) and this anger induces some people to punish free riders; that is, punishment reflects punitive sentiments. Given that punishment is expected, self-regarding people now have an incentive to cooperate. Modern lawful societies channel punitive

sentiments into laws and a formal, institutionalized sanctioning system, which provide incentives to cooperate.

What matters from the point of view of a self-regarding individual is the expected cost of free riding. The presence of peer punishment might make cooperation worthwhile, but so can incentives provided by other mechanisms. For example, O’Gorman, Henrich, and Van Vugt (2009) and Baldassarri and Grossman (2011) studied centralized punishment by one group member and found it quite effective. Centralized punishment can even be effective if it is not deterrent (Engel 2013). Falkinger and coauthors (2000) showed that an exogenously given tax-subsidy mechanism induces people to cooperate in line with theoretical predictions about how the incentives should work. Another line of research, dating back to a seminal paper by Toshio Yamagishi (1986), showed that people are also willing to contribute to a “punishment fund” (think of funding law enforcement through people’s taxes) to punish lowest contributors. Comparing (the evolution of) peer punishment and pool punishment has triggered theoretical investigations (Sigmund et al. 2010) and is also an important topic of experimental research (e.g., Traulsen, Röhl, and Milinski 2012; Zhang et al. 2014).

In the remainder of this section, I discuss briefly two mechanisms other than punishment that have also proved effective in supporting cooperation. The mechanisms I will consider are rewards; and indirect reciprocity and the role of a good reputation.

*Rewards.* Because punishment is successful in increasing cooperation (under perfect observability), an intuitive question is whether rewards can also sustain cooperation. Punishment, whenever it is used, has the disadvantage that it is costly for the punisher as well as for the punished person (i.e., punishment is inefficient because resources are destroyed). Rewards do not have this disadvantage. They might be costly too for the rewarding person, but if the benefits of the reward at least cover the costs, rewards are not inefficient.

Most experiments model rewards analogously to punishment: after group members have made their contributions, they are informed about each contribution made and can then allocate reward points to the target group member. One reward point costs 1 money unit and the rewarded group member then gets, depending on the experiment, one or more money units as an additional payment. The results suggest that this mechanism can also stimulate contributions, in particular if the rewarded individual receives more than what it costs to reward (Sefton, Shupp, and Walker 2007; Rand et al. 2009; Sutter, Haigner, and Kocher 2010). For example, in experiments comparable to Gächter, Renner, and Sefton (2008) summarized in figure 2.4, Rand and coauthors (2009) showed that achieved cooperation levels were as high as those under punishment.

It is important to notice that there is a fundamental asymmetry between punishments and rewards: rewards have to be used to be effective, whereas under punishment a credible threat can suffice (figure 2.4). Thus, punishment can be very cheap, whereas rewards will be costly. Moreover, in a context of law enforcement rewards are typically the exception and threats of punishment the norm.

*Indirect reciprocity and reputation.* Humans keenly care about their reputation. Why? The mechanism of indirect reciprocity (Nowak and Sigmund 1998) provides

an important likely channel. People not only help those who helped them (direct reciprocity) but might also help those who helped others. Thus, if one has a reputation of helping others, one might receive more help as well, and it pays to be a cooperator. Experimental evidence supports this theoretical argument (e.g., Milinski, Semmann, and Krambeck 2002). Relatedly, people's concerns to be held in good esteem can stimulate prosocial behavior (e.g., Ariely, Bracha, and Meier 2009). Evidence for the success of reputation-based incentives is not restricted to the lab. For example, a recent field experiment showed that a concern for good reputation can help in energy conservation, which is an important public good in the real world (Yoeli et al. 2013).

## 6 RULE OF LAW AND SELF-GOVERNANCE OF SOCIAL DILEMMA PROBLEMS

---

The research I have presented so far has mostly been conducted in a few Western societies, such as the United States, Britain, and Switzerland. How representative are these societies when making claims or inferences about human nature? According to an influential study by Henrich, Heine, and Norenzayan (2010) there is substantial heterogeneity in human behavior across the many societies on this planet that makes the Western societies look like outliers. Does this also hold for the behavioral patterns reported in this chapter?

The existence of strong positive and negative reciprocity has been shown in many societies around the world (Henrich et al. 2005; Henrich et al. 2006). Herrmann, Thöni, and Gächter (2008) conducted a series of public goods experiments without and with punishment in fifteen quite different large-scale societies around the world (such as the United States, Turkey, China, Saudi Arabia, and England; see figure 2.5). They uncovered three important findings relevant for the present topic. First, without punishment cooperation breaks down everywhere (figure 2.2). Second, with punishment, it turns out that people punish free riders very similarly across the fifteen societies. In stark contrast, there is substantial cross-societal variation in antisocial punishment, that is, punishment of people who contributed to the public good by people who contributed less than the group member they punish. Third, there is a very large variation in cooperation levels achieved and, due to antisocial punishment, cooperation does not always raise contributions compared to the condition without punishment. Figure 2.5A illustrates the cooperation levels achieved and their relation to antisocial punishment.

The results by Herrmann, Thöni, and Gächter (2008) provide us with an important caveat on the power of punishment to stimulate prosocial cooperation. Punishment only increases cooperation if it is targeted towards free riders exclusively; antisocial punishment is a huge impediment to successful cooperation. Relatedly, punishment can only stimulate cooperation if it does not trigger counterpunishment (e.g., Nikiforakis 2008).

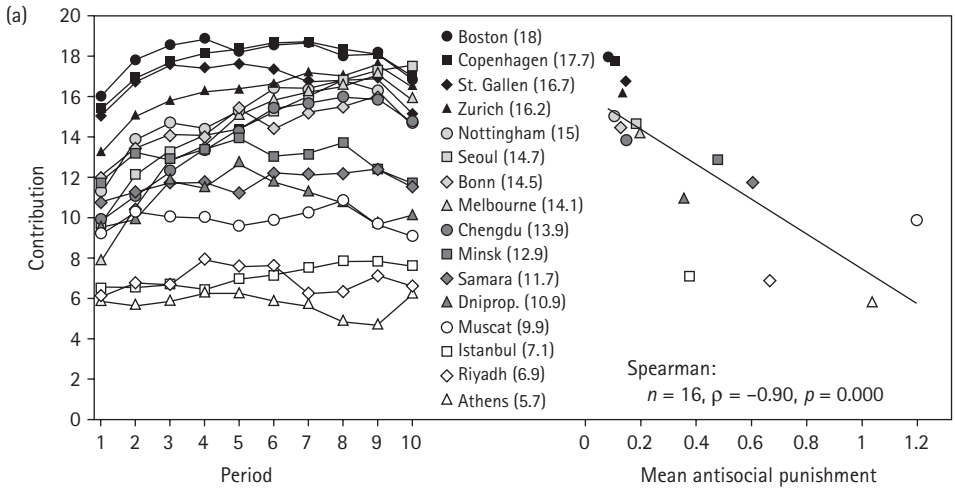


FIGURE 2.5A Cross-Societal Cooperation and Antisocial Punishment

Data source: Herrmann, Thöni, and Gächter (2008).

Figures source: Gächter and Thöni. 2011.

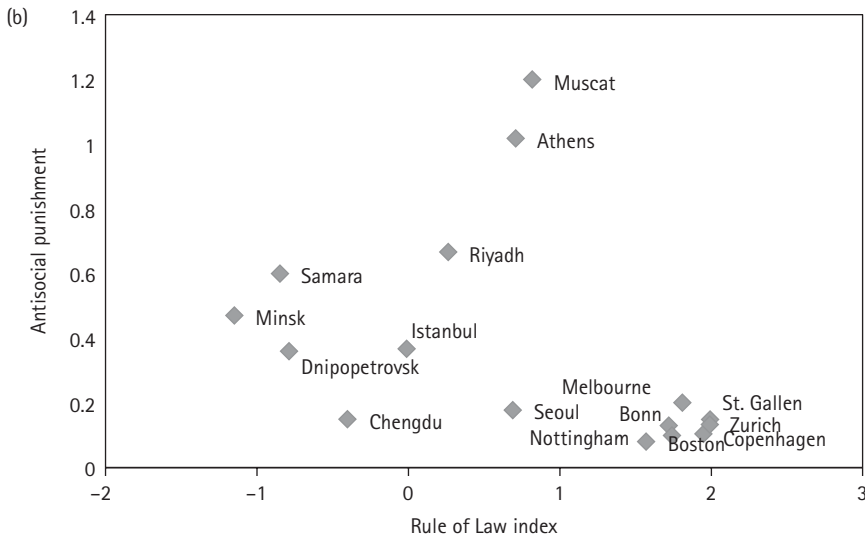


FIGURE 2.5B The Stronger Is the Rule of Law the Lower Is Antisocial Punishment

Data source: Herrmann, Thöni, and Gächter (2008); own illustration.



The Herrmann, Thöni, and Gächter (2008) study reveals another relevant finding, namely that the severity of antisocial punishment in a society is linked to the Rule of Law in that society. The Rule of Law indicator is a governance indicator developed by the World Bank to measure how well private and government contracts can be enforced in courts, whether the legal system and police are perceived as being fair, how important the black market and organized crime are, and so on (see Herrmann et al. 2008 for details). Figure 2.5B illustrates how the Rule of Law is linked to antisocial punishment observed in a given society.

The results are quite striking. The Western societies all have a high Rule of Law index value, and there is also very little antisocial punishment observed in these societies. The variation in antisocial punishment increases substantially once the Rule of Law index falls below 1 (the theoretical range is between  $-2.5$  and  $+2.5$ ).

The significance of this finding is twofold. First, the fact that experimentally measured behavior is correlated to societal measures suggests that the societal background has an influence on behavior. The studies by Henrich and coauthors (2005), Henrich and coauthors (2006) and Henrich, Ensminger, and coauthors (2010) suggest such an influence based on the organization of the small-scale societies where they conducted their research. The Herrmann, Thöni, and Gächter (2008) findings show that societal background also matters for developed, large-scale societies. Second, and more importantly for present purposes, the negative correlation of antisocial punishment and the quality of the Rule of Law in a society suggests that a high-quality law enforcement system (which can be interpreted as a high degree of institutionalized cooperation) will also limit antisocial punishment and thereby an important inhibitor of voluntary cooperation. Good institutions make for good self-governance of people who manage to cooperate with one another and who limit punishment to those who fail to cooperate.

## 7 SUMMARY AND CONCLUDING REMARKS

---

In this chapter I have provided from two decades of behavioral economics research evidence that, rather than being selfish as is assumed in the *homo economicus* paradigm, many people are strong reciprocators, who punish wrongdoing and reward kind acts. However, a sizable minority of people is best characterized as selfish. My main focus has been on determinants of social order, which I have construed as a social dilemma where individual incentives are not aligned with collective benefits.

I have argued that from the perspective of the behavioral science of cooperation, and in particular strong reciprocity, social order has three important determinants: the strength of internalized norms of prosocial behavior, the behavior of other people, and the threat of punishment or the presence of other incentives to curb selfishness. Looking at the many results in synthesis suggests the following big picture: many people are motivated by character virtues such as honesty and trustworthiness; they think



that free riding is morally blameworthy; they feel guilty if it turns out that others contributed more to the public good than them; they are angry at the free riders; and they experience some warm glow by contributing to the public good. However, all research shows that people are also very strongly looking at the behavior of others to determine their behavior. Since a sizable number of people are free riders and even many conditional cooperators have a selfish bias, cooperation in randomly assembled groups is inherently fragile. Cooperation can only be sustained under the strong requirement that only highly cooperatively inclined people are matched and able to exclude free rider types. Under more realistic conditions, stable prosocial cooperation requires some incentives, most notably punishment, where often a credible threat suffices to keep free riding at bay.

Notice that the three determinants of social order are also linked. If norms are strong and induce many people to cooperate, then the psychology of conditional cooperation will induce many people to cooperate as well. However, because a sizable minority of people is not motivated by normative considerations but only by own gain, norms appear a rather weak determinant of social order because conditional cooperators will only cooperate if others do as well. In other words, the psychology of conditional cooperation appears to be the stronger behavioral force than normative considerations and, as a consequence, cooperation will be fragile. This conclusion follows from three separate observations I recorded in this chapter: (1) character virtues and normative considerations including feelings of guilt if others behave more cooperatively matter for many people (figure 2.1); (2) conditional cooperation is an important motivation for the average person (figure 2.3) and (3) cooperation nevertheless almost inevitably breaks down if not backed up by incentives (figure 2.2). Punishment (like other incentives) has the dual advantage that it induces the free rider types to cooperate and thereby convinces the conditional cooperators to maintain their cooperation.

I conclude this chapter with some remarks on future research. Of the three determinants of social order the first determinant (the role of norms, moral judgments, and emotions such as guilt for cooperation) is the least well understood determinant of cooperation. More research is necessary to understand people's normative consideration and to what extent this influences their behavior. With regard to the second determinant (conditional cooperation) an important open question is gaining a complete picture of proximate mechanisms that determine conditional cooperation including gauging their relative importance. The third determinant (punishment and other incentives) is the best-understood determinant. Open questions are finding explanations for antisocial punishment and how antisocial punishment is causally related to the Rule of Law (figure 2.5). A further underresearched topic is the role of institutional punishment in successful cooperation, in particular in comparison with peer punishment and when considering the role of errors and imperfect observability. Finally, an important topic for future research is to understand how exactly the three determinants are linked and how the three determinants work in naturally occurring settings.

## ACKNOWLEDGEMENT

---

Support under the European Research Council grant ERC-AdG 295707 COOPERATION is gratefully acknowledged. I thank Doron Teichman and Eyal Zamir for their comments and the Phoolan Devi Institute and Department of Social Psychology at Free University Amsterdam for their hospitality while working on this chapter.

## REFERENCES

---

- Ambrus, A., and B. Greiner. 2012. Imperfect public monitoring with costly punishment: An experimental study. *American Economic Review* 102:3317–32.
- Anderson, C. M., and L. Putterman. 2006. Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior* 54:1–24.
- Andreoni, J. 1990. Impure altruism and donations to public-goods—a theory of warm-glow giving. *Economic Journal* 100:464–77.
- Ariely, D., A. Bracha, and S. Meier. 2009. Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review* 99:544–55.
- Axelrod, R. 1984. *The evolution of cooperation*. New York: Basic Books.
- Baldassarri, D., and G. Grossman. 2011. Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences* 108:11023–27.
- Balliet, D., L. B. Mulder, and P. A. M. Van Lange. 2011. Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin* 137:594–615.
- Balliet, D., C. Parks, and J. Joireman. 2009. Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations* 12:533–47.
- Baron-Cohen, S. 2011. *Zero degrees of empathy: A new theory of human cruelty*. London: Allen Lane.
- Berg, J., J. Dickhaut, and K. McCabe. 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10:122–42.
- Bolton, G. E., and A. Ockenfels. 2000. ERC: A theory of equity, reciprocity, and competition. *American Economic Review* 90:166–93.
- Bornstein, G., and O. Weisel. 2010. Punishment, cooperation, and cheater detection in “noisy” social exchange. *Games* 1:18–33.
- Bowles, S., and H. Gintis. 2011. *A cooperative species: Human reciprocity and its evolution*. Princeton, NJ: Princeton University Press.
- Boyd, R., and P. J. Richerson. 1988. The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology* 132:337–56.
- Brekke, K. A., G. Kipperberg, and K. Nyborg. 2010. Social interaction in responsibility ascription: The case of household recycling. *Land Economics* 86:766–84.
- Buckholtz, J. W., and R. Marois. 2012. The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience* 15:655–61.
- Carpenter, J. 2004. When in Rome: Conformity and the provision of public goods. *Journal of Socio-Economics* 33:395–408.

- Casari, M., and L. Luini. 2012. Peer punishment in teams: Expressive or instrumental choice? *Experimental Economics* 15:241–59.
- Charness, G., and P. Kuhn. 2011. Lab labor: What can labor economists learn from the lab? In *Handbook of labor economics*, O. Ashenfelter and D. Card. Amsterdam: Elsevier, 229–330.
- Charness, G., and M. Rabin. 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117:817–69.
- Chaudhuri, A. 2011. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics* 14:47–83.
- Chaudhuri, A., and T. Paichayontvijit. 2006. Conditional cooperation and voluntary contributions to a public good. *Economics Bulletin* 3:1–15.
- Cooter, R. 2000. Do good laws make good citizens? An economic analysis of internalized norms. *Virginia Law Review* 86:1577–601.
- Cubitt, R., M. Drouvelis, and S. Gächter. 2011. Framing and free riding: Emotional responses and punishment in social dilemma games. *Experimental Economics* 14:254–72.
- Cubitt, R., M. Drouvelis, S. Gächter, and R. Kabalin. 2011. Moral judgments in social dilemmas: How bad is free riding? *Journal of Public Economics* 95:253–64.
- Dufwenberg, M., S. Gächter, and H. Hennig-Schmidt. 2011. The framing of games and the psychology of play. *Games and Economic Behavior* 73:459–78.
- Dufwenberg, M., and G. Kirchsteiger. 2004. A theory of sequential reciprocity. *Games and Economic Behavior* 47:268–98.
- Eckel, C., and P. Grossman. 1996. Altruism in anonymous dictator games. *Games and Economic Behavior* 16:181–91.
- Egas, M., and A. Riedl. 2008. The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B—Biological Sciences* 275:871–78.
- Ehrhart, K.-M., and C. Keser. 1999. Mobility and cooperation: On the run. Working Paper No. 99s-24, CIRANO Montreal.
- Ellickson, R. 1991. *Order without law: How neighbors settle disputes*. Cambridge: Harvard University Press
- Engel, C. 2011. Dictator games: A meta study. *Experimental Economics* 14:583–610.
- Engel, C. 2013. Deterrence by imperfect sanctions—a public good experiment. Preprints of the Max Planck Institute for Research on Collective Goods Bonn 2013/9.
- Engelmann, D., and M. Strobel. 2004. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review* 94:857–69.
- Falk, A., E. Fehr, and U. Fischbacher. 2003. On the nature of fair behavior. *Economic Inquiry* 41:20–26.
- Falk, A., E. Fehr, and U. Fischbacher. 2005. Driving forces behind informal sanctions. *Econometrica* 73:2017–30.
- Falk, A., and U. Fischbacher. 2006. A theory of reciprocity. *Games and Economic Behavior* 54:293–315.
- Falk, A., and J. Heckman. 2009. Lab experiments are a major source of knowledge in the social sciences. *Science* 326:535–38.
- Falkinger, J., E. Fehr, S. Gächter, and R. Winter-Ebmer. 2000. A simple mechanism for the efficient provision of public goods: Experimental evidence. *American Economic Review* 90:247–64.
- Fehr, E., and U. Fischbacher. 2004. Social norms and human cooperation. *Trends in Cognitive Sciences* 8:185–90.

- Fehr, E., U. Fischbacher, and S. Gächter. 2002. Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature* 13:1–25.
- Fehr, E., and S. Gächter. 2000. Cooperation and punishment in public goods experiments. *American Economic Review* 90:980–94.
- Fehr, E., and S. Gächter. 2002. Altruistic punishment in humans. *Nature* 415:137–40.
- Fehr, E., G. Kirchsteiger, and A. Riedl. 1993. Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics* 108:437–59.
- Fehr, E., and K. M. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114:817–68.
- Fehr, E., and K. M. Schmidt. 2006. The economics of fairness, reciprocity and altruism—experimental evidence and new theories. In *Handbook of the economics of giving, altruism and reciprocity*, S.-C. Kolm and J. M. Ythier. Amsterdam: Elsevier B.V., 615–91.
- Fischbacher, U., and S. Gächter. 2010. Social preferences, beliefs, and the dynamics of free riding in public good experiments. *American Economic Review* 100:541–56.
- Fischbacher, U., S. Gächter, and E. Fehr. 2001. Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters* 71:397–404.
- Fischbacher, U., S. Gächter, and S. Quercia. 2012. The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology* 33:897–913.
- Forsythe, R., J. Horowitz, N. E. Savin, and M. Sefton. 1994. Fairness in simple bargaining experiments. *Games and Economic Behavior* 6:347–69.
- Frey, B. S., and S. Meier. 2004. Social comparisons and pro-social behavior. Testing “conditional cooperation” in a field experiment. *American Economic Review* 94:1717–22.
- Frey, B. S., and B. Torgler. 2007. Tax morale and conditional cooperation. *Journal of Comparative Economics* 35:136–59.
- Friedman, D., and S. Sunder. 1994. *Experimental methods: A primer for economists*. Cambridge: Cambridge University Press.
- Gächter, S., and B. Herrmann. 2009. Reciprocity, culture, and human cooperation: Previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B—Biological Sciences* 364:791–806.
- Gächter, S., E. Renner, and M. Sefton. 2008. The long-run benefits of punishment. *Science* 322:1510.
- Gächter, S., and C. Thöni. 2005. Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association* 3:303–14.
- Gächter, S., and C. Thöni. 2011. Micromotives, microstructure and macrobehavior: The case of voluntary cooperation. *Journal of Mathematical Sociology* 35:26–65.
- Galbiati, R., and P. Vertova. 2008. Obligations and cooperative behaviour in public good games. *Games and Economic Behavior* 64:146–70.
- Gintis, H. 2000. Strong reciprocity and human sociality. *Journal of Theoretical Biology* 206:169–79.
- Gintis, H., S. Bowles, R. Boyd, and E. Fehr, Eds. 2005. *Moral sentiments and material interests. The foundations of cooperation in economic life*. Cambridge: MIT Press.
- Glimcher, P. W., C. F. Camerer, E. Fehr, and R. A. Poldrack, Eds. 2009. *Neuroeconomics. Decision making and the brain*. Amsterdam: Elsevier.
- Gneezy, U. 2005. Deception: The role of consequences. *American Economic Review* 95:384–94.
- Grechenig, C., A. Nicklisch, and C. Thöni. 2010. Punishment despite reasonable doubt—a public goods experiment with uncertainty over contributions. *Journal of Empirical Legal Studies* 7:847–67.

- Grujić, J., B. Eke, A. Cabrales, J. A. Cuesta, and A. Sánchez. 2012. Three is a crowd in iterated prisoner's dilemmas: Experimental evidence on reciprocal behavior. *Scientific Reports* 2:10.1038/srep00638.
- Gürerk, Ö., B. Irlenbusch, and B. Rockenbach. 2006. The competitive advantage of sanctioning institutions. *Science* 312:108–11.
- Güth, W., R. Schmittberger, and B. Schwarze. 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3:367–88.
- Haidt, J. 2003. The moral emotions. In *Handbook of affective sciences*, R. J. Davidson, K. R. Scherer, and H. H. Goldsmith. Oxford: Oxford University Press, 852–70.
- Henrich, J., R. Boyd, S. Bowles, C. F. Camerer, E. Fehr, and H. Gintis. 2004. *Foundations of human sociality. Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford: Oxford University Press.
- Henrich, J., R. Boyd, S. Bowles, C. F. Camerer, E. Fehr, H. Gintis, R. McElreath, M. Alvard, A. Barr, J. Ensminger, N. Henrich, K. Hill, F. Gil-White, M. Gurven, F. W. Marlowe, J. Q. Patton, and D. Tracer. 2005. “Economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences* 28:795–855.
- Henrich, J., J. Ensminger, R. McElreath, A. Barr, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker. 2010. Markets, religion, community size, and the evolution of fairness and punishment. *Science* 327:1480–84.
- Henrich, J., S. J. Heine, and A. Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33:61–83.
- Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J.-C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, T. David, and J. Ziker. 2006. Costly punishment across human societies. *Science* 312:1767–70.
- Herrmann, B., and C. Thöni. 2009. Measuring conditional cooperation: A replication study in Russia. *Experimental Economics* 12:87–92.
- Herrmann, B., C. Thöni, and S. Gächter. 2008. Antisocial punishment across societies. *Science* 319:1362–67.
- Hume, D. 1987 [1777]. *Essays: Moral, political and literary*. Indianapolis: Liberty Fund.
- Johnson, N. D., and A. A. Mislin. 2011. Trust games: A meta-analysis. *Journal of Economic Psychology* 32:865–89.
- Kahan, D. 2003. The logic of reciprocity: Trust, collective action, and law. *Michigan Law Review* 102:71–103.
- Kahneman, D., and A. Tversky, Eds. 2000. *Choices, values, and frames*. Cambridge: Cambridge University Press.
- Kirchgässner, G. 2008. *Homo oeconomicus: The economic model of behaviour and its applications in economics and other social sciences*. New York: Springer.
- Kirchler, E. 2007. *The economic psychology of tax behaviour*. Cambridge: Cambridge University Press.
- Kocher, M. G., T. Cherry, S. Kroll, R. J. Netzer, and M. Sutter. 2008. Conditional cooperation on three continents. *Economics Letters* 101:175–78.
- Lewinsohn-Zamir, D. 1998. Consumer preferences, citizen preferences, and the provision of public goods. *Yale Law Journal* 108:377–406.
- López-Pérez, R., and E. Spiegelman. 2013. Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics* 16:233–47.

- Martinsson, P., N. Pham-Khanh, and C. Villegas-Palacio. 2013. Conditional cooperation and disclosure in developing countries. *Journal of Economic Psychology* 34:148–55.
- Milinski, M., D. Semmann, and H. J. Krambeck. 2002. Reputation helps solve the “tragedy of the commons” *Nature* 415:424–26.
- Nikiforakis, N. 2008. Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics* 92:91–112.
- Nikiforakis, N., and H. Normann. 2008. A comparative statics analysis of punishment in public goods experiments. *Experimental Economics* 11:358–69.
- Nowak, M. A., and K. Sigmund. 1998. Evolution of indirect reciprocity by image scoring. *Nature* 393:573–77.
- O’Gorman, R., J. Henrich, and M. Van Vugt. 2009. Constraining free riding in public goods games: Designated solitary punishers can sustain human cooperation. *Proceedings of the Royal Society B-Biological Sciences* 276:323–29.
- Olson, M. 1965. *The logic of collective action*. Cambridge: Harvard University Press.
- Oosterbeek, H., R. Sloof, and G. van de Kuilen. 2004. Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics* 7:171–88.
- Ostrom, E. 1990. *Governing the commons. The evolution of institutions for collective action*. Cambridge: Cambridge University Press.
- Ostrom, E., J. M. Walker, and R. Gardner. 1992. Covenants with and without a sword—self-governance is possible. *American Political Science Review* 86:404–17.
- Posner, E. A. 2000. *Law and social norms*. Cambridge: Harvard University Press.
- Pruckner, G. J., and R. Sausgruber. 2013. Honesty on the streets: A field study on newspaper purchasing. *Journal of the European Economic Association* 11:661–79.
- Rabin, M. 1993. Incorporating fairness into game-theory and economics. *American Economic Review* 83:1281–302.
- Rand, D. G., A. Dreber, T. Ellingsen, D. Fudenberg, and M. A. Nowak. 2009. Positive interactions promote public cooperation. *Science* 325:1272–75.
- Rand, D. G., and M. A. Nowak. 2013. Human cooperation. *Trends in Cognitive Sciences* 17:413–25.
- Rapoport, A., and A. M. Chammah. 1965. *Prisoners’ dilemma. A study in conflict and cooperation*. Ann Arbor: University of Michigan Press.
- Riker, W. H., and P. C. Ordeshook. 1968. A theory of the calculus of voting. *American Political Science Review* 62:25–42.
- Rustagi, D., S. Engel, and M. Kosfeld. 2010. Conditional cooperation and costly monitoring explain success in forest commons management. *Science* 330:961–65.
- Schwartz, S. H. 1977. Normative influences on altruism. In *Advances in experimental social psychology*, L. Berkowitz. New York: Academic Press, 221–79.
- Sefton, M., R. Shupp, and J. M. Walker. 2007. The effect of rewards and sanctions in provision of public goods. *Economic Inquiry* 45:671–90.
- Shang, J., and R. Croson. 2009. A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *Economic Journal* 119:1422–39.
- Shinada, M., and T. Yamagishi. 2007. Punishing free riders: Direct and indirect promotion of cooperation. *Evolution and Human Behavior* 28:330–39.
- Sigmund, K. 2007. Punish or perish? Retaliation and collaboration among humans. *Trends in Ecology and Evolution* 22:593–600.



- Sigmund, K., H. De Silva, A. Traulsen, and C. Hauert. 2010. Social learning promotes institutions for governing the commons. *Nature* 466:861–63.
- Stigler, G. J. 1981. Economics or ethics? In *Tanner lectures on human values*, S. McMurrin. Cambridge: Cambridge University Press.
- Sunstein, C., Ed. 2000. *Behavioral law and economics*. Cambridge: Cambridge University Press.
- Sutter, M., S. Haigner, and M. G. Kocher. 2010. Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies* 77:1540–66.
- Thöni, C., J.-R. Tyran, and E. Wengström. 2012. Microfoundations of social capital. *Journal of Public Economics* 96:635–43.
- Traulsen, A., T. Röhl, and M. Milinski. 2012. An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proceedings of the Royal Society B: Biological Sciences* 279:3716–21.
- Traxler, C. 2010. Social norms and conditional cooperative taxpayers. *European Journal of Political Economy* 26:89–103.
- Van Lange, P. A. M., D. Balliet, C. D. Parks, and M. Van Vugt. 2014. *Social dilemmas. The psychology of human cooperation*. Oxford: Oxford University Press.
- Volk, S., C. Thöni, and W. Ruigrok. 2012. Temporal stability and psychological foundations of cooperation preferences. *Journal of Economic Behavior & Organization* 81:664–76.
- Wilkinson, N., and M. Klaes. 2012. *An introduction to behavioral economics*. 2nd edition. Basingstoke: Palgrave Macmillan.
- Yamagishi, T. 1986. The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology* 51:110–16.
- Yamagishi, T., N. Mifune, Y. Li, M. Shinada, H. Hashimoto, Y. Horita, A. Miura, K. Inukai, S. Tanida, T. Kiyonari, H. Takagishi, and D. Simunovic. 2013. Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Organizational Behavior and Human Decision Processes* 120:260–71.
- Yoeli, E., M. Hoffman, D. G. Rand, and M. A. Nowak. 2013. Powering up with indirect reciprocity in a large-scale field experiment. *Proceedings of the National Academy of Sciences* 110:10424–29.
- Zhang, B., C. Li, H. Silva, P. Bednarik, and K. Sigmund. 2014. The evolution of sanctioning institutions: An experimental approach to the social contract. *Experimental Economics* 17:285–303.

## CHAPTER 3

---

# MORAL JUDGMENT

---

JONATHAN BARON

## 1 HISTORY OF MORAL JUDGMENT RESEARCH

---

MORAL judgment is at the heart of legal judgment. It affects the interpretation of laws in specific cases, the creation of laws, and public support for them. Moral philosophy and legal philosophy have large areas of overlap, to the point where they may often be confused. Psychology has studied moral judgment since the nineteenth century, as have other social sciences.

### 1.1 Piaget and Kohlberg

Perhaps the most important historical influence today is the tradition begun by Jean Piaget (1932). Piaget was interested in what we now call the cognitive development of children, but he called it “genetic epistemology”. The term “genetic” referred to genesis, that is, development, rather than inheritance.

His method was to interview children about hypothetical situations and ask follow up questions. For example, he asked them about the game of marbles, as an analogy to social rules. He asked who made the rules, and whether they could be changed. Younger children thought that the rules could not be changed.

Piaget’s view was that (to paraphrase thousands of pages in a few words) children develop through an inevitable process of interaction with their environment. For moral judgment, this was the social environment.

This view may be compared to two alternatives on either side. One is that people are a blank slate and will learn from their environment, whatever it is. Piaget would dispute this in part because he would argue that the kinds of interactions are predetermined by the nature of the situation and are thus not so flexible. It is therefore possible to make generalizations about development that hold across cultures.



The other alternative view is that moral judgments are literally genetic in the biological sense, determined by “cognitive modules” that are the result of natural selection (e.g., Haidt & Joseph 2004; Mikhail 2011). Such an account is always possible, but critics note that humans have had very little time for natural selection to work on specific moral rules. We might think that such selection might require language to exist already. Although it is easy to tell a story about how any particular moral principle could have been adaptive at some point in human prehistory, it is so difficult to test such assertions that they are almost unscientific. Moreover, such accounts are not needed. I will ignore evolutionary accounts in this chapter. My own view is broadly Piagetian.

Piaget’s cognitive theory was extended by Lawrence Kohlberg (1963; Colby & Kohlberg 1987), who used the same interview methods, but much more systematically, with objective scoring systems and rules. A typical Kohlberg dilemma concerned whether Heinz should steal a drug to save his wife’s life, when the inventor of the drug refused to sell it for what Heinz could afford to pay.

Kohlberg’s theory had three levels, with two stages in each. Levels 1 and 2 were characterized by a failure to make a certain distinction in moral judgment. At level 1, children fail to distinguish morality from self-interest. When asked whether some behavior is wrong, they reply by noting whether it is likely to be punished, or come back to haunt the actor in some other way: “Heinz should save his wife so she could cook for him”; or “Heinz should not steal the drug because he would be put in jail.” At the level 2, children fail to distinguish morality and social convention. They appeal to social norms or law (civil law or religious law) to justify their judgments: “A good husband would steal it”; or “Theft is a sin”. At level 3, people distinguish between convention and morality and can use moral judgments to criticize law or social norms, for example, by arguing that it is sometimes justified to break a law that is too strict.

Research using Kohlberg’s scoring methods found that development through the stages was affected by opportunities to reflect on moral issues. More interestingly, although Piaget generally felt that the kinds of cognitive development he studied were complete by the time of adolescence, Kohlberg and his colleagues found that development to the highest level occurred in a minority of adults in developed countries, and in almost no adults in countries without extensive educational systems. Thus, for Kohlberg, much of the world did not think that laws and social norms are subject to reasoned criticism.

Elliot Turiel (1983) argued that Kohlberg’s interview methods gave too little credit to young children. When he posed direct yes/no questions to them, he was able to show that even most five-year-olds could distinguish morality from convention. For example, children said that it was wrong to push another child out of a swing, and that it was wrong for a boy to wear a dress to school. But, when asked whether it would still be wrong if everyone thought it was OK, they said yes for pushing but no for wearing a dress. Turiel’s distinction has come to be called “authority independence” and is now taken as a way of distinguishing true moral judgments from judgments based on convention (e.g., Royzman, Leeman, & Baron 2009).

Turiel's results illustrate a general principle of cognitive development (Flavell 1971). Development of a distinction such as that between morality and convention (or length and number) occurs gradually, possibly over many years. During the period of development, children will pass "easier" tests of knowledge of the distinction but fail harder ones. Turiel's yes/no measures were easier to pass than Kohlberg's interview measures.

Yet it turns out that Kohlberg might have been correct about the sequence and about cultural differences. Haidt, Koller, and Dias (1993) used Turiel's methods to assess moral judgment in groups different in age and (expected) educational level in Brazil and the United States. (Expected educational level was the typical education that children would receive later.) Some items were intended to be both harmless and offensive, such as "A family's dog was killed by a car in front of their house. They had heard that dog meat was delicious, so they cut up the dog's body and ate it." In general, Brazilian children and American children with low education or educational expectation did not make the moral/conventional distinction, even with yes/no measures like those used by Turiel.

A special contribution of the line of research that began with Piaget is its emphasis on the relation between moral judgment and general cognitive development, such as the idea of making more relevant distinctions with greater cognitive maturity. An important contribution of Kohlberg is the suggestion that most adults are not fully cognitively mature. The claim of incomplete development is perhaps an optimistic version of the more general claim that adults show cognitive biases and that these biases affect moral judgment along with other judgments and decisions, as discussed later.

## 1.2 Equity Theory and Economic Games

Another historical source of current research on moral judgment comes from the study of moral *behavior*, specifically, of motives to be fair. According to *equity theory* (Walster, Walster, & Berscheid 1978), people desire to see that outcomes are just or equitable. Because of this desire, people try to restore equity when outcomes are inequitable. They therefore reward those who have been underrewarded and punish those who have been overrewarded. They do this even when they must sacrifice some of their own reward. For example, when subjects in experiments are overpaid, they tend to work harder in the next experiment, as if they want to make their overpayment fair.

More interestingly, when people cannot restore equity, they try to deceive themselves into believing that the winners deserved to win and the losers deserved to lose. For example, they will ascribe negative personality traits to a person who is badly treated, but only when they cannot help the victim (Lerner & Simmons 1966). Thus, the subjects "derogated the victim" only when they could not restore equity themselves. Lerner and Simmons suggest that we desire to believe that the world is orderly and fair, lest we fear that we ourselves will become victims of unfairness. This "just world" hypothesis has been supported by many other experiments (Lerner & Miller 1978).

The field of experimental economics has rediscovered motives for fairness and has put them in a different context, that of economic theory (chapter 2 by Gächter in this volume). In a *dictator game*, one subject—the dictator—is asked to divide an amount of money between herself and another subject. The other subject has no choice about whether to accept the offer or not; hence, the first subject is the “dictator.” In one version of this game, Kahneman, Knetsch, and Thaler (1986a) asked each student in a class to divide \$20 with an anonymous other student in the same class. There were only two options: an equal division of \$10 each, and an unequal division of \$18 for the “dictator” and \$2 for the other. Out of the whole group of subjects, 76% divided the money equally. In a second part of the experiment, some subjects made a second decision about whether to pay a dollar in order to punish a student who had chosen the \$18 option, leaving a third student (not the subject) with \$2. The punishment deprived the unfair student of additional money. Most student given this choice were willing to give up \$1 for this purpose. In sum, subjects are, once again, willing to sacrifice their narrow self-interest for the sake of fairness, specifically for the punishment of prior unfairness toward others.

Another case in which subjects sacrifice narrow self-interest in order to punish unfairness is the *ultimatum game*. Suppose you are told that you and a student in a different class have a chance to divide up \$10. The other student, the offerer, will offer you some part of the \$10. If you, the receiver, accept the offer, you will get the amount offered, and the offerer will get the rest. If you reject the offer, you will each get nothing. Would you accept an offer of \$5? \$2? 1 cent? Offers much below \$5 are often rejected. The receiver prefers to see both subjects get nothing rather than tolerate such unfairness. The receiver therefore is willing to sacrifice in order to punish unfairness. (This is the negative side of fairness motivation, the desire to hurt others, even at one’s own expense, in order to restore equality.) Notice that if receivers were concerned only with their self-interest, they would accept an offer of 1 cent, and the offerers, knowing that, would offer only 1 cent. Most offerers offer \$5, or only a little less—perhaps out of a desire for fairness, perhaps out of fear of rejection, perhaps for both reasons (Thaler 1988).

In sum, people desire to see fairness. They often try to bring about fairness, even if they must sacrifice to bring it about. When they cannot bring it about, they often try to deceive themselves into thinking that things are fair. Although most of these experiments are about behavior, and this chapter is about judgment, it should be noted that judgments what is right are generally consistent with what people do (e.g., Eichenberger & Oberholzer-Gee 1998).

### 1.3 Intuition and Experiments in Philosophy

Philosophers, legal scholars, judges, and moral theologians have for centuries used their own intuitive judgments as the basis for drawing conclusions about morality,

although they did not always admit that they were doing this. Noam Chomsky (1957) inspired more reflection about the role of intuitions when he explicitly set out to make his own intuitions the basis for developing a theory of linguistic syntax. He limited the relevant intuitions to those concerning whether a sentence was grammatical or not, so that, famously, “Colorless green ideas sleep furiously” was counted as a grammatical sentence, but it would not be grammatical without the verb “sleep.” In contrast to many philosophers, he argued that a theorist must be willing to reject occasional intuitions that were inconsistent with an otherwise highly coherent theory. John Rawls (1971) applied Chomsky’s approach to morality, arguing that the moral theorist must strive for “reflective equilibrium” between theory and data—the data being the intuitions, some of which might have to be declared incorrect.

Chomsky was after an essentially psychological account of the human capacity for language. He was not concerned with an ideal theory of what language should be, but rather with a theory of what it was. But Rawls is usually interpreted as claiming that he was after a truly normative theory of morality, a theory of what it should be. Yet his methods were those that would yield a theory of moral judgment without telling us anything about whether our theory-consistent judgments *should* be followed in any given case. And in fact it is difficult to find in *A Theory of Justice* (Rawls 1971) a clear statement that the result was anything but a psychological theory of our concept of justice. (That said, Rawls did argue that our natural conception of justice limited the possible political and legal systems that could be stable, but this argument seems to me to be an open empirical question.)

Despite such concerns, many philosophers have assumed that Rawls’s approach is a legitimate method for discovering the moral truth, in whatever sense it can be said to exist. In an important essay, Greene (2007) summarizes the case against the use of intuition in any way. Greene points out that intuitive judgments are known to be distorted by various biases, in clear cases when there is a normative right answer, so we should not trust these intuitions as a path to discovering the right answer itself. And the theory that could develop from our own intuition might simply reflect common cultural and historical influences, such as Christianity (Singer 2005).

Recently a number of philosophers have started to do experiments because they think that these experiments will help solve philosophical problems (Knobe & Nichols 2008). The experiments are much like those described in the next subsection. Although such experiments may be better in some ways than one philosopher relying on just her own intuitions (but worse in other ways, such as the difficulty of communicating the exact question to nonphilosophers), this approach is controversial for many of the same reasons that any use of intuition is controversial.

Much less controversial is the use of experimental methods by philosophers not so much as input to philosophical analysis but rather as illustrations (e.g., Bicchieri 2006), often contributing to psychology itself.

## 1.4 Moral Heuristics and Biases

Greene's (2007) view leads naturally to the possibility of approaching moral judgment within the heuristics-and-biases framework described in chapter 1 by Baron in this volume. In order to apply this framework, we need a normative theory of moral judgment, a standard for how we define the right answer to hypothetical moral questions, or the right choices in economic games. Then we look for biases and ultimately ask how we might improve judgments or decisions according to the standard.

This approach was (to my knowledge) first applied explicitly by Ritov and Baron (1990) and Spranca, Minsk, and Baron (1991), and defended more generally by Baron (1994). The studies in question were concerned with the distinction between acts and omissions. For example, Spranca, Minsk, and Baron asked subjects to compare two endings to a story, in which Ellen had decided to lie to the police to protect her friend. In one ending, Ellen lied. In the other, she was about to lie when she realized that the police had decided incorrectly that her friend was innocent, and she fails to correct their false belief. Subjects were asked to compare Ellen's morality in the two endings. Although many subjects thought she was equally wrong in both cases—given that her intent was the same and the outcome, which she could easily have prevented, was the same—others thought that she was more wrong in the first cases, where she acted.

Ritov and Baron (1990) asked about giving a vaccine that would prevent a disease but would cause side effects that were equivalent to the disease in some smaller number of people. Failing to give the vaccine was a harmful omission, which was more harmful than the act of vaccinating, yet many subjects favored not vaccinating. This problem was presented as a personal decision and as a policy decision, with essentially the same results. The policy decision could be considered moral. The authors did not distinguish moral and personal decisions, but simply assumed that they were both subject to the same bias, which they called "omission bias."

## 2 UTILITARIANISM AS A NORMATIVE THEORY

---

In saying that the result demonstrates a bias, we implicitly assumed a normative model, which in this case implied that the optimal response was to minimize the overall harm. This conclusion is implied by a utilitarian theory, but it may be disputed by other theories, for example, theories that disparage harming some people as means to help other people. Those who are harmed by the vaccine can be seen as means in this way.

In the rest of this chapter, I shall assume a utilitarian view as normative. I have defended this theory elsewhere (Baron 1993a, 1996, 2006, 2008), and others have

defended it with great depth and sophistication (Hare 1981). But I shall make a few comments on the general strategy.

One is that researchers may have to take sides in philosophical disputes, in order to make a claim of bias or systematic error. We cannot, for example, always rely on experimental methods such as “framing effects” to avoid such commitments (Baron 1994). Framing effects exist when different descriptions of the same situation yield different judgments. Yet the criterion of the “same situation” often begs philosophical questions; specifically, it is usually understood as meaning that the consequences are the same. But consequentialism itself is a controversial moral theory, so it must be defended explicitly if any consequentialist theories are to be taken as normative.<sup>1</sup>

Other investigators have taken the same approach while at the same time not committing themselves as to the normative status of utilitarianism. They regard it as an interesting model for psychological purposes. Utilitarianism says that the best option is the one with the best overall consequences, and it is of interest to know why people sometimes judge some other option to be better, even if the answer is that they are following the truly correct moral theory. I have argued that this noncommittal approach, in combination with one other assumption, has substantial value. The additional assumption is that people’s judgments about what is morally best lead to choices that, to a large extent, accomplish what the people intend them to accomplish. Thus, if we want to understand why things do not always turn out for the best, one possible answer is that people are trying to do something else, and to some extent succeeding. It may be that the “something else” is to follow a better moral theory, but then at least we learn the price we pay, in terms of consequences, for morality.<sup>2</sup>

Utilitarianism is a theory about decision-making. It tells us what to choose, when we are choosing between options that affect many people. Sometimes the utilitarian prescription will conflict with the prescription of what decision best serves our self-interest. These are, I think, two different questions.<sup>3</sup> Similarly, the simple form of the theory does not distinguish such concepts as “permitted,” “forbidden,” “duty,” and “supererogation” (beyond the call of duty). These make sense in the law, but not in the simplest form of utilitarian theory, which answers only one question: when faced with

<sup>1</sup> Consequentialism is the claim that only consequences matter in evaluating options. Utilitarianism is a form of consequentialism in which it is also claimed that we can evaluate overall consequences by adding up effects on different people. The vaccination example requires utilitarianism, since it involves adding up effects over different numbers of people. Consequentialist theories, and different versions of utilitarianism, may differ in how consequences are evaluated. Modern forms of utilitarianism, such as that of Hare (1981), generally think of utility in terms of something like degree of satisfaction of “preferences” (a term that requires further definition).

<sup>2</sup> The assumption that we are likely to get what we try to get is needed to block another argument, often made for forms of rule utilitarianism, which is that the best consequences, by a utilitarian standard, are actually achieved when people try to do something else. To me this alternative seems more plausible when applied to interpersonal interactions, where things like loyalty and love may be more important than anything else, but highly implausible in the context of law and public policy, which is, as it happens, the context of many experiments in this field.

<sup>3</sup> But see Baron & Szymanska (2010) for a discussion of one way to think about this conflict.

options for a choice that affects several people, which one is best? Of course, utilitarianism can be applied to decisions about the law itself, as Bentham did.

What is called “welfare economics” overlaps with utilitarianism. This form is often used, to varying degrees, in discussions of the economic analysis of law (e.g., Shavell 2004). The main characteristic of this approach is that it allows the possibility of interpersonal comparisons of “welfare,” which can be understood as the sort of utility that various forms of utilitarianism seek to maximize. Some forms of welfare economics depart from the assumption that utilities can be added up without any transformation (e.g., Adler 2012), but that assumption typically has very little effect on the analysis of laws themselves (Shavell 2004).

## 2.1 Relation to Expected Utility

Utilitarianism is closely related to expected-utility theory, the standard normative model for individual decisions under uncertainty. As a result, parallels are possible between research on biases in decisions that affect others and biases in decisions that affect only the self.

Specifically, when a group of people all face the same decision, then the two models clearly dictate the same choice. For example, suppose each of 1,000 people faces a 2% probability of some disease without vaccination, but the vaccine causes an equally serious disease with a probability of 1%. The best decision for each person is to get the vaccine. If everyone gets the vaccine, then we expect that 100 will get the disease caused by the vaccine, instead of 200 getting the other disease. Suppose that a policymaker could decide whether to give the vaccine to everyone or not. The vaccine would maximize each person’s utility. But omission bias would lead to the opposite choice. For an individual, we would usually not call this a moral decision or a moral principle. Nobody else is involved. But a policymaker might find it immoral to harm others through action. She might try to follow a rule such as “Do no harm (through action).” Omission bias has been observed in decision for both self and others (e.g., Ritov & Baron 1990).

## 2.2 Conflicts with Intuition

Empirical researchers study conflicts between moral theories and moral intuition. If only because people have different intuitions, no moral theory can capture them all. But conflicts between intuitive moral judgments and utilitarian theory are of interest not only to empirical researchers but also to philosophers, who debate the proper role of intuition in evaluation and construction of normative moral theories (e.g., Rawls 1971; Hare 1981; Greene 2007).

Utilitarian philosophers have offered a number of suggestions about why utilitarianism often conflicts with moral intuition, particularly Hare (1981, especially ch. 8). Examples of cases that inspire such intuitions are whether a mother should save her



son in an airplane crash, in preference to a world-famous surgeon who would go on to save many others; and whether an emergency-room doctor should kill a vagrant who wanders in to get warm, in order to save five other people from immediate death by transplanting the vagrant's organs.

Hare (and others) argue that our moral intuitions are not designed for such unlikely situations and that they serve us well most of the time. The utilitarian answer may be correct, but the intuitions may be too strong for us to accept it. Moreover, such strong intuitions may be part of a good utilitarian character. It may be very difficult to induce people to feel strongly about morality yet at the same time be willing to make exceptions for such unusual special cases. And the apparent utilitarian answer may not be the best utilitarian answer, which may take into account the effect of choices on others.

Last but not least, when we face such unusual situations in real life, and we think we know the utilitarian right answer, we should, if we are good utilitarians, also consider the probability that our perception is incorrect, which can sometimes be quite high. Most adulterers and terrorists think that their behavior serves the greater good (that their spouses will not find out and thus will not suffer harm, or that a massacre of noncombatants will bring on utopia), yet they might temper their judgments if they considered that practically every adulterer and terrorist before them has thought the same thing, and in hindsight most of them have been wrong. People faced with unusual hypothetical situations might carry over such caution without fully knowing where their caution comes from.

Such arguments as these serve as warnings to those researchers who would want to label apparent nonutilitarian intuitions as biases, relative to the utilitarian normative theory itself. People who give the wrong utilitarian answer about whether to give a vaccine, or whether to punish the innocent in order to deter others, may be responding on the basis of rules that can be justified by utilitarianism itself. What they may (or may not) lack is understanding of that justification.

### 3 PROTECTED (SACRED) VALUES (PVs)

---

I now turn to a discussion of possible biases in moral judgment, from the perspective of utilitarianism as a normative model. The first example is the existence of moral rules that are taken as absolute, not to be violated no matter what the cost. For example, people may think that it is always wrong to kill innocent people, no matter what benefits result from doing so (including, perhaps, the benefit of saving others from death). When people explicitly affirm such rules, even when they admit that the benefits exceed the costs of violations, then this affirmation violates utilitarianism. Absolute values as such do not necessarily violate it: people might believe that one consideration has infinite utility relative to others. In this case, they will say it is impossible for the benefits to exceed the costs. While such judgments do not violate utilitarianism, they



often seem odd, and they sometimes conflict with other judgments, for example, when two absolute rules conflict with each other.

Baron and Spranca (1997) called such attitudes “protected values” (PVs) because they were protected from trade-offs with other values. Many of these values involve morally questioned behaviors, such as cloning people for reproduction or destroying the natural environment. Other researchers have used the term “sacred values” for what seems to be the same phenomenon, although they measure it differently (Fiske & Tetlock 1997; Tetlock, Lerner, & Peterson 1996). Roth (2007) has also used the term “repugnant transactions” for moral prohibitions on transactions such as a live donor selling a kidney.

Rules of this sort, if they were taken seriously, would cause great difficulty for evaluation of public policies through measurement of utility, because they would amount to infinite utilities and would thus allow one person to determine a decision for everyone—unless someone else had a conflicting rule, in which case the choice could not be determined. People who endorse more than one such rule—and many people endorse several—could find themselves in a similar dilemma.

It appears that absolute rules are often overgeneralizations (chapter 1). When people are asked to try to think of counterexamples, cases in which the benefits would be great enough to justify taking the prohibited action, they can usually do so, and they change their mind about whether the rule is absolute (Baron & Leshner 2000). Thus, PVs could be explained psychologically as the result of failure to think critically about rules. A rule against killing sounds good, until one tries to think of counterexamples. Of course, over time, people may become committed to some rules, so that they resist counterarguments. Thus, rules may be maintained by “my-side bias” (Baron 2008 and chapter 1 in this volume).

PVs are closely related to, and correlate with, other types of values or goals, which express themselves in other ways. In particular, they are related to *moralistic goals*. These goals or values are those that people want others to follow, regardless of whether the others endorse the same goals and regardless of whether the consequences are, on the whole, worse as a result. Baron (2003) found that people endorsed moralistic goals for banning actions like the following:

- Testing a fetus for IQ genes and aborting it if its expected IQ is below average
- Cloning someone with desired traits so that these may be passed on, such as an athletic champion or brilliant scientist
- Modifying the genes of an embryo so that, when it is born, it will have a higher IQ
- Giving a drug (with no side effects) to enhance school performance of normal children

In many cases (22% of examples like these), subjects (recruited on the World Wide Web) would ban these actions even if the consequences of allowing the actions were better on the whole than the consequences of banning them, if the subjects could imagine that the consequences might be better, and if “almost everyone in a nation thought that the behavior should be allowed.” In sum, they were willing to impose their moral(istic)

principles on others, whatever the consequences, and whatever the others wanted. Moralistic values, taken seriously, are protected from trade-offs with values involving consequences, so they are likely to be thought of as PVs.

## 4 OMISSION, ACTION, AND RELATED ISSUES

---

Omission bias, a second example of a moral bias, is the tendency to judge acts that are harmful (relative to the alternative option) as worse than omissions that are equally harmful (relative to the alternative) or even more harmful (as in the vaccination case) (Baron & Ritov 1994). In any given case, some people display this bias and others do not.

Omission bias is related to issues of public controversy, such as whether active euthanasia should be allowed. Most countries (and most states of the United States) now allow passive euthanasia, the withholding of standard medical treatment for those who are judged to be no worse off dead than alive, but active euthanasia and assisted suicide are banned in most places even for those who wish to die. Opponents of active euthanasia can, of course, find other arguments against it than the fact that it is “active.” But it is possible that these arguments would not be seen as so compelling if the distinction between acts and omissions were not made.

The bias is related to other issues in the law. Although it might occasionally be possible to punish people for extremely harmful omissions of actions that are easy to take, very few jurisdictions have laws against such omissions (Feldbrugge 1966). Omission bias also seems to explain the fact that the actual burden of proof on plaintiffs in lawsuits is substantially higher than implied by the official standard of “preponderance of the evidence,” which implies a probability of .5 for the correctness of the plaintiff’s claim: triers regard accepting the plaintiff’s claim as an action, and denying it as an omission, and they are much more afraid of erroneous acts than of erroneous omissions (Zamir & Ritov 2012).

### 4.1 Trolley Problems

Psychological interest in omission bias was inspired not just by real cases such as euthanasia or resistance to vaccination but also by examples devised by philosophers. The most famous of these is the trolley problem of Foot (1978), later developed by other philosophers and psychologists. The basic question is whether one should switch a runaway trolley from a track where it is headed for five people (who will be killed) to a different track where it is headed for one person. Different studies ask whether one should switch the trolley, whether it is permissible to do it, and whether doing it is morally required. The “should” question is the one that most directly assesses utilitarian thinking, since utilitarians may regard the other questions as fundamentally legal

questions, not moral ones, given that utilitarianism in its simplest form is just about how to determine which option is morally best.

Since Spranca, Minsk, and Baron (1991), many other papers have examined responses to the trolley problem and variants of it (e.g., Greene et al. 2009, and citations therein). The most heavily studied issue involves a contrast between the basic case (switching) and one in which the only way to stop the trolley is to push a fat man off a bridge, so that the fat man will be hit and killed, but stop the trolley and save the five others. The result of this and several other variants is that people consider harm to be worse when it involves physical contact with the victim, as I discuss shortly.

## 4.2 Determinants of Omission Bias

Omission bias has been studied extensively, and we are in a position to draw some conclusions about why it happens. In particular, research has identified several moderators of the effects, that is, manipulations that affect its frequency, and correlates of its presence.

### 4.2.1 *Perceived Causality*

An early and much replicated result is that omission bias is correlated with differences between acts and omissions in perceived causality. When people say that harms of action are worse than equivalent harms of omission, they also tend to think that the causal relation between the person and the harm is greater in the action (Spranca, Minsk, & Baron 1991; Baron & Ritov 2009a).

Two concepts of causality compete here and elsewhere. One is “but for” (*sine qua non*) causality. The other is physical causality. We say that person P causes outcome O, in this sense, if O was affected by P’s choice.<sup>4</sup> But-for causality does not distinguish acts and omissions. Thus, in tort law, people can be held liable for omissions, when they are in a position of responsibility.

In physical causality, P’s behavior is linked to O through a series of physical events, each of which presumably follows some physical principle. When people distinguish acts and omissions morally, they seem to be basing moral judgments on this kind of causality.

But-for causality may be harder for young children to learn. The appreciation of it requires evaluation of counterfactuals. Although children as young as five can distinguish cases of causality by omission from noncausality (Schleifer, Shultz, & Lefebvre-Pinard 1983), the distinction appears to be weak. The cases were very simple. For example, a store owner did not put salt on the ice in front of his store. In one condition, a customer slipped on the ice and was injured. In the control condition, the customer fell before reaching the ice, and was injured.

<sup>4</sup> If P’s choice affected O probabilistically, we say that it was a partial cause or a contributing cause.

#### 4.2.2 *Physical Proximity, Contact, and Personal Force*

Most people regard it as worse to “push a man off of a footbridge and in front of a train in order to cause the man to fall and be hit by the train, thereby slowing it and saving five people ahead on the tracks” than to “pull a lever that redirects a trolley onto a side track in order to save five people ahead on the main track if, as a side-effect, pulling the lever drops a man off a footbridge and in front of the train on the side tracks, where he will be hit” (Cushman, Young, & Hauser 2006). Although this result seems to be due to physical proximity, Greene and coauthors (2009) showed that, in one situation, this sort of case depends on “personal force” more than on contact itself, and proximity alone had no effect at all. Personal force means that the force that directly affects the victim “is generated by the agent’s muscles, as when one pushes another with one’s hands or with a rigid object.”

#### 4.2.3 *Protected Values*

PVs are also related to omission bias. PVs are absolute rules. Such rules would, to put it mildly, be difficult to follow if they were neutral between acts and omissions (as I noted earlier). If you think that abortion is absolutely wrong no matter what the benefit, it is easy for you to take no action that causes an abortion, but it is extremely difficult to avoid omissions that lead to abortion. If you tried to do this, you would be morally obliged to spend your time doing little else but trying to stop abortions from happening. If you had two such PVs against omission, then you would be in real trouble. People seem to recognize this logical asymmetry, and, as a result, they endorse PVs against action more than PVs against omission (Baron & Ritov 2009a).

When people have PVs against some action, then that action usually shows strong omission bias. If asked, for example, whether it is right to kill one person in order to save 5, 50, or 500 others, the numbers don’t matter. A person with a true PV against active killing will always say no, and many subjects do exactly this (Ritov & Baron 1999; Baron & Ritov 2009a).

### 4.3 Related Biases

Several biases in moral judgment are related to omission bias. In many cases, the biases are confounded, so that it is impossible to say which one is present. I know of no attempt to distinguish individual differences in these. It might turn out that some common factor accounts for many of them, such as perceived causality.

#### 4.3.1 *Indirectness and the Double Effect*

The indirectness bias is illustrated in the doctrine of the double effect. For example, when a mother’s life is threatened by a pregnancy, some Catholic hospitals will permit a hysterectomy to save the mother, but they will not permit an abortion. The fetus dies in either case, but, in the case of the hysterectomy (which of course leaves the mother

unable to bear another child), the killing is seen as an indirect by-product (Bennett 1966; Kuhse 1987). In the abortion, however, the death of the fetus is the means to save the mother, so the fetus is being harmed directly. The indirectness bias is shown in the following scenario (Royzman & Baron 2002):

A new viral disease is spreading rapidly in a region of Africa. Left alone, it will kill 100,000 people out of 1,000,000 in the region. X, a public health official, has two ways to prevent this. Both will stop the spread of the virus and prevent all these deaths:

A. Give all 1,000,000 a shot that makes them immune to the first disease. The shot will also cause, as a side effect, a second disease that will kill 100 people.

B. Give all 1,000,000 a shot that gives them another disease, which is incompatible with the first disease. The second disease will kill 100 people.

Most subjects thought that option A was better, because the deaths are a side effect rather than part of the mechanism of the main effect. Indirectness bias is also related to perceived causality (Baron & Ritov 2009a).

#### 4.3.2 *Agent Relativity*

Agent relativity illustrated in the following scenario used by Baron and Miller (2000). X is one of ten people who could save someone's life by donating bone marrow (a painful but relatively risk-free procedure) to Y. Is X's obligation to donate greater when X is Y's cousin than when X and Y are unrelated? Many people think so. Utilitarians even think so, if they think that family cohesion is a good thing that should be promoted for other reasons. Now consider Z, who is unrelated to X or Y. X, the potential donor, asks Z's advice about whether to donate, and Z knows that X will probably follow the advice offered. Does Z have a greater obligation to advise donation when X and Y are cousins than when X and Y are unrelated? A utilitarian who answered yes to the first question would have to answer yes to this one. After all, it is promoting family loyalty that is at issue, and it doesn't matter whose family it is (without knowing more details, of course). An *agent relative* response, however, would say that only Y needs to worry about family obligations. The obligation is relative to the agent. It differs from person to person. Miller and Baron found no evidence for agent relativity in any of their subjects (who were Indian as well as American). However, many philosophers argue that some obligations are agent relative in this way. (See McNaughton & Rawling 1991, for a review.)

Omission bias is agent relative when the harm from omission is the result of someone else's action. In the classic case of shooting one prisoner to save ten from being shot by a horrible dictator, the choice of not shooting is obviously agent relative, because shooting will happen anyway. This is not a pure test of agent relativity, though, because the two options also differ in doing something versus doing nothing. (The agent should be required to shoot at a target to indicate to the dictator that he will not shoot the prisoner.)

### 4.3.3 *Naturalism*

Naturalism is the bias toward nature. It is also related to omission bias, because “nature” often defines the default situation, the result of inaction, as in the case of the vaccination, where the disease can be assumed to be natural. Of course, the result of omission is not always natural, as in the case of the dictator just described. (See Rudski et al. 2011, for a recent review.)

## 5 THE PSYCHOLOGICAL BASIS OF MORAL BIASES

---

The results for omission bias are perhaps the clearest illustration of the idea of moral heuristics and biases, so at this point I will step back and discuss some general issues in the psychology of these biases. Then I will review additional results.

### 5.1 Two-Systems Theory

Greene (e.g., 2007) has proposed a theory of the conflict between utilitarianism and deontology in omission-bias dilemmas. The theory relies on the idea of two “systems” of cognitive processing (see chapter 1 by Baron in this volume). System 1 is fast, automatic, and effortless. System 2 requires effortful thinking. Greene also proposes that System 1 is influenced by emotional responses, more than System 2. Thus, in a dilemma such as the fat-man version of the trolley problem, people have an immediate, automatic emotional response to the idea of pushing a man to his death, and this leads them to want to say that it would be wrong to do so. Then, some people will reflect before they make this response, using System 2, and decide that they would not want to let five others die through their inaction.

Several lines of evidence support this theory, yet questions can be raised about each one.

First, response times (RTs) for “personal” dilemmas like the one at issue are longer. Baron and coauthors (2012) argue that this result can be explained in terms of conflict. When choices are difficult, so that the subject is as likely to respond yes as no, RT is long. By the two-system theory, at this point, it should still be longer for yes than no responses, because yes responses require an extra step. This result is not found.

Second, cognitive interference slows down RT to utilitarian responses but not deontological responses (Greene et al. 2008). The interference, however, involved arithmetic, and it was necessary to process the numbers to give the utilitarian response. Also, the deontological response could be based on a less thorough reading of the dilemma,

considering only the type of action to be taken and not the numbers. Similar issues arise in other studies.

Third, Suter and Hertwig (2011) have claimed that instructing people to go fast or slow affected their responses. Yet they found this only for a few selected dilemmas, and a reanalysis of their data shows no overall effect. Moreover, Gürçay and Baron (in preparation) have failed to find such an effect in two studies.

Fourth, utilitarian responding correlates with cognitive reflection as a trait (e.g., Paxton, Ungar, & Greene 2011). Cognitive reflection is measured by a test of that name (chapter 1 by Baron in this volume), which consists of three short arithmetic problems with “obvious” answers that turn out to be incorrect, thus apparently requiring correction of a System 1 response by System 2. This is a more interesting result, but a correlation like this does not imply that correction is involved in the moral judgment task, or even in the arithmetic test itself. A person who is not distracted by the trick answers in the arithmetic test might just adopt an attitude of using System 2 from the outset, analyzing each problem without even being tempted to take a guess at the answer. Similarly, in moral judgment, people may set out to look at all the information, including side effects of doing nothing, before even making a tentative judgment. More generally, the correlation could result from individual differences in reflection-impulsivity (Baron, Badgio, & Gaskins 1986), a measure of cognitive style concerned with the relative preference for accuracy (reflection) versus speed (impulsivity).

This kind of account is not far from the two-system account, but it does not assume any sequential effects involving suppressing an early response by a late one, so it is thus consistent with the results discussed so far in this section.

It is also possible that other factors could affect the observed correlations. For example, people who do well on the arithmetic problems might have had some type of education that exposes them to utilitarian arguments.

## 5.2 The Role of Emotion

The evidence on the role of emotion is also strong. Much of it concerns individual differences, some of them the result of brain damage. Damage to regions of the brain that involve emotional responses is correlated with more utilitarian judgments (e.g., Moretto et al. 2010, who find evidence that the effects are correlated with actual emotional responses). Psychopathy is a trait associated with immoral behavior, is also associated with blunted emotional reactions, and is also correlated with utilitarian judgments in the usual dilemmas, as are other similar traits (e.g., Bartels & Pizarro 2011). Interestingly, the tendency to feel anger, unlike other emotions, may be positively correlated with utilitarian responses (Choe & Min 2011).

Several attempts have been made to manipulate emotion and show an effect on moral judgment. To my knowledge, none of these attempts has convincingly showed that such a direct effect of emotion on the content of moral judgment (as distinct from their intensity) can occur. By “content” I mean whether the judgment favors a utilitarian or



deontological solution, while “intensity” refers to the judged appropriate punishment for transgressions.

Some of the problems with this research concern the nature of the experimental manipulations. For example, some studies manipulate disgust by using the term itself, which has a direct moral meaning in English as well as its emotional meaning (Nabi 2002). Others manipulate disgust by presenting something disgusting, such as a smell. Subjects in such experiments might be thinking that the experimenter is immoral to allow such things in the laboratory. Thus, here too it is difficult to separate the emotion itself from a moral judgment. Yet other studies use manipulations that supposedly reduce emotion by asking people to be dispassionate, but the wording of such instructions also encourages subjects to question their initial judgment cognitively. Finally, even the studies that show results with these problems unsolved turn out to be difficult to replicate. Several unreported studies in which I have been involved have repeatedly found no effects of manipulation of emotion. In sum, my current belief is that a direct causal effect of emotion on moral judgment about hypothetical cases will not be found, but I could be proved wrong by subsequent research.<sup>5</sup>

How, then, can we explain the correlations between emotional propensity and moral judgment? One way is to reverse the direction of cause and effect. In particular, when people are presented with a hypothetical dilemma such as the fat man, they often respond with an intuitive, cognitive, condemnation of the act of pushing, and this moral judgment evokes the emotional response. The condemnation is absent in brain-damaged patients and psychopaths. Another possible account (consistent with the first) is that emotion is involved in the learning of principles of moral judgment, but once they are learned the principles continue to be applied even if the emotion is absent. But in normal people it is still present. Psychopaths, because they lack the emotional response, do not so easily learn that causing harm through personal force is wrong.

Notice that the difficulty of finding a causal effect of emotion on judgment causes problems for a currently popular theory of moral judgment (Haidt 2001), which holds that moral judgments are mainly the result of System 1 reasoning evoked by an emotional response and that System 2 reasoning is almost always rationalization rather than reasoning that is capable of leading to a change. Nor can this view be saved by removing the emotion, so long as it assumes that the System 1 response is almost always deontological. Such a proposal is inconsistent with the observation of large individual differences (e.g., Baron et al. 2012). The evidence indicates that reasoning (as subjects perceive it) is involved in producing answers to moral dilemmas (Bucciarelli, Khemlani, & Johnson-Laird 2008).

The association of emotion with deontological responding is also suspect. This seems to be the result of the attention given to a few cases in the psychology literature, cases

<sup>5</sup> I am leaving out details here, because this situation is obviously in flux. My point is that we cannot accept at face value the extant studies that claim to show direct causal effects of emotion on moral judgment.



such as the fat man and other similar dilemmas in which the utilitarian response is pitted against something awful and disturbing, if only to show that people are sometimes not utilitarian reasoners. It is not at all clear that real-world dilemmas are predominantly of this type, as opposed to the opposite type in which the utilitarian response is the one supported by emotion, particularly emotions such as empathy, while the deontological response is the result of rigid application of a rule, such as Kant's famous example in which he argues that it would be wrong to lie in order to save someone's life (Baron 2011a; Kahane et al. 2012). Very few people today would agree with Kant on this point, but people often apply rules of their religion or culture even when they must force themselves, against obvious human passions, to do so.<sup>6</sup> One example might be following a rule against abortion, even when it saves the mother's life and when the fetus would die anyway.

### 5.3 Developmental and Historical Origins of Deontology

Discussions of the role of intuition often suggest that it arises from special evolved capacities, often described as modules, analogous to those that we seem to have for learning language, learning to recognize faces, and other specialized tasks (Mikhail 2011). I do think that human morality is something special, which goes way beyond what other animals have, but before proposing such specialized evolutionary applications—especially given the very short time that natural selection had to work on our species before we took reproduction largely out of its hands—I think it is safer to try to find more general and less specialized ways of accounting for our specialness (Baron 2011b). One of course is the existence of language itself, which gives us enormous capacity to communicate culture to each other. Another is the parallel development of the capacity to be influenced by each other, which Herbert Simon called “docility” (Knudsen 2003). Perhaps related to that is also our capacity to empathize. Once we have language plus docility, then culture becomes very powerful, and we are likely to find more answers by looking at the evolution of culture than at the biological evolution of humans.

## 6 BIASES IN HELPING OTHERS

---

Returning to the discussion of moral biases, away from utilitarianism as a normative model, we find that several biases have been identified in choices that involve helping others.

<sup>6</sup> See Kahane et al. 2012, for a similar argument, with some results.

## 6.1 Psychophysical Numbing and the Value of Lives

One of the classic demonstrations in the field of decision-making is the Asian Disease problem of Tversky and Kahneman (1981), in which it is shown that people tend to be risk averse when it comes to saving lives (preferring to save 200 for sure than a one-third chance of saving 600) but risk seeking when it comes to preventing deaths (preferring a program in which 600 will die with probability two-thirds over a program in which 400 will die for sure). An implication of this result that went without notice for many years is that people do not value other people's lives equally, even when the others are completely unknown and presumably drawn from the same population.

Slovic (2007) realized the importance of this phenomenon, and cited extensive other literature making the same point. In particular, people seem to have declining marginal disutility for other people's deaths. That is, the death of 1 million people is not perceived as 1,000 times as bad as the death of 1,000. In the extreme, Josef Stalin is quoted as saying, "When one man dies, it is a tragedy. When thousands die it's statistics." Slovic pointed out that this result is a consequence of a general psychological principle, diminishing sensitivity. When we judge differences, we judge them as smaller when they are farther away from where we are, that is, from our reference point (nobody dead, or nobody saved). Yet, argues Slovic, this is an error with great consequences, because it causes us to ignore large human tragedies, such as genocides, that could be prevented or reduced at a relatively small cost per life saved. He calls it "psychophysical numbing."

## 6.2 Identified Victim

A single life, as Stalin noted, is the smallest unit of all and thus acquires special significance. Some charities tell potential donors that their money will go to help a specific child, in contrast to those who admittedly take all the donations and put them in a big pot of money, which they then distribute so as to help a great many children. Research has found that focusing on a single, identified "victim" promotes more altruism (e.g., Small & Loewenstein 2005; Kogut & Ritov 2005; Slovic 2007 reviews the literature).

## 6.3 Heuristics and Biases in Charity

Other biases have been suggested for charitable donations (Baron & Szymanska 2010): People tend to minimize waste even when minimizing waste reduces total benefit; in some situations the more wasteful charity does more good despite the waste, yet people want to avoid it. People attend to average benefit per dollar rather than to the marginal benefit of additional contributions, thus possibly failing to contribute enough to relatively poorly funded organizations doing good work, where a few extra dollars could matter a lot. People diversify their contributions, as if they were

investments, despite the utilitarian optimal strategy of donating all to what seems to be the best bet (which is especially optimal because it leads to organizations spending less money mailing out appeals). Some people prefer voluntary charity to taxation and government assistance, which they view as “forced charity.” And people are biased toward causes in their own nation, even when the same amount of money can do more good elsewhere. (As always in discussing moral judgment, “people” means “some people.” There are always individual differences.)<sup>7</sup>

## 7 FAIRNESS IN DISTRIBUTION OF GOODS AND BADS

---

Many issues in morality and law concern fairness and distribution. Obvious examples are the distribution of income that results from taxation policies and the distribution of goods such as access to healthcare or even organs for transplantation. But we can also include the distribution of bads, such as punishment. Thus, tort law and criminal law can be included under this general rubric.

### 7.1 Equality Heuristics

People sometimes want to divide resources equally even when the overall benefit from unequal division is greater. Several studies have presented subjects with allocation dilemmas of the following sort (Ubel, Baron, & Asch 2001; Ubel & Loewenstein 1996): Two groups of 100 people each are waiting for liver transplants. Members of group A have a 70% chance of survival after transplantation, and members of group B have a 30% chance. How should 100 livers—the total supply—be allocated between the two groups? The simple utilitarian answer is all to group A, but typically less than 20% of the subjects will choose this allocation. People want to give some livers to group B, even if less than half. Many want to give half. Some of this effect is the result of not knowing how to maximize. When subjects are asked what would maximize survival, some say that the allocation they chose would do so. But others make an explicit

<sup>7</sup> The utilitarian normative theory for charity is somewhat unclear. Many writers have argued that we should give to charity until the harm from our loss is equal to the benefit from our contribution, a strategy that would make us all quite poor. Baron and Szymanska argue for a more limited theory, which assumes that altruism (willingness to sacrifice self-interest for others) is limited. Within the constraints of limited altruism, utilitarianism then implies that we should do the most good for whatever we are willing to contribute, which is approximately fixed and not part of our decision. Utilitarianism thus becomes (roughly) a theory for decision-making, with motivation being an external fact, rather than a theory of what our motivation should be.

distinction between maximizing survival and being fair. They are willing to trade lives for fairness.

The equality principle can conflict with itself, because different quantities can be equated, such as income and costs of a business. Harris and Joyce (1980) told subjects about various situations in which a group of partners had opened a business (for example, selling plants at a flea market). The partners took turns operating the business, so different amounts of income were generated while each partner was in control, and different costs were incurred. (Costs for one partner were extremely high because of an accident.) When subjects were asked how the *profits* should be divided among the partners, many subjects favored equal division. When subjects were asked how they would divide up the *expenses*, they tended to divide *these* equally, even though they understood that the resulting division of profits would be unequal (because each partner would keep what was earned during her turn, after paying her share of the expenses). Judgments of fairness seem to exhibit framing effects (chapter 1 Baron in this volume).

## 7.2 Penalties and Punishment

### 7.2.1 Torts

Tort damages serve two functions: deterrence and compensation.<sup>8</sup> Compensation is like insurance. Indeed, its function could be replaced by insurance (including the possibility of social insurance, that is, paid for from taxes rather than individual premiums). The utilitarian justification of insurance is declining marginal utility: if you lose something that can be replaced by paying money (i.e., a pecuniary damage), like a house, then your utility for money is suddenly greater, and we can maximize overall utility by taking money from others, that is, those who pay house insurance premiums, and giving it to you. In the case of simple torts with pecuniary damages, the optimal penalty is equal to the replacement cost of the damage (with various simplifying assumptions—see Shavell 2004). If you know you have to pay the cost of damages, you will take sufficient care to avoid them. But for nonpecuniary damages such as death, compensation is impossible or at least not the same as the penalty required for deterrence.

In order to separate the two functions, Baron and Ritov (1993) asked subjects to assess penalties and compensation separately for victims of birth-control pills and vaccines (in cases involving no clear negligence). We found evidence that many people do not think about deterrence. For example, in one case, subjects were told that a higher penalty would make the company and others like it try harder to make safer products. In an adjacent case, a higher penalty would make the company more likely

<sup>8</sup> Sometimes these functions are separated, as in New Zealand. Separation makes it easier to optimize both functions.

to stop making the product, leaving only less safe products on the market. Most subjects, including a group of judges, assigned the same penalties in both of these cases. When the utilitarian principle of deterrence was suggested as a way of distinguishing the two cases, some subjects thought it was a good idea, and others thought it was morally wrong and that penalties should be proportional to the harm without looking ahead to future effects.

We also found (as had Baron 1993b) that compensation awards were affected by such normatively irrelevant factors as whether an injury was caused by people or by nature.

A second bias in judgments of penalties is that people seem to want to make injurers undo the harm they did, even when some other penalty would benefit others more. For example, Baron, Gowda, and Kunreuther (1993) found that subjects preferred to have companies clean up their own waste, even if the waste threatened no one, rather than spend the same amount of money cleaning up the much more dangerous waste of a defunct company. Ordinarily, it is easiest for people to undo their own harm, but this principle may be overgeneralized. (See also Beattie & Baron 1995.)

Both of these biases can lead to worse consequences in some cases, although much of the time the heuristics that lead to them probably generate the best consequences. These results, then, might also be the result of overgeneralization of otherwise useful heuristics.

### 7.2.2 *Crimes: Deterrence versus Retribution*

Paul Robinson and John Darley (e.g., 2007) have argued that judgments of appropriate criminal penalties fit an intuitive concept of retribution rather than deterrence. They also argue that support for the law would be weakened if the law went against these intuitions.

Baron and Ritov (2009b) following Sunstein, Schkade, and Kahneman (2000), tested one implication of the deterrence model, which is that the severity of the penalty should be higher when the probability of apprehension is lower, other things being equal. Most subjects did not take probability into account spontaneously, and many subjects thought that it should not matter, because the punishment should match the severity of the offense.

Such a concept of retribution may be seen as a heuristic. In many cases, retribution and deterrence agree. Yet such a heuristic is difficult to justify in terms of making the task easier. It is not obviously easier to match the length of a prison term to the severity of injury in an assault than to think about whether the prison term would deter other would-be offenders.

## 7.3 The Morality of Citizenship

The moral judgment literature is mostly concerned with individual decisions. Such decisions are especially important in the case of moral behavior (see generally

chapter 9 by Feldman in this volume; chapter 10 by Bilz and Nadler in this volume) when self-interest is pitted against doing what is right, as in all the classic cases of conflict of interest. But moral judgment is also important for decisions that do not usually involve much of this sort of conflict, namely, the decisions of citizens, when they vote or participate politically in other ways. Although a citizen's judgments, when expressed, have little effect on that citizen, the collective decisions of all citizens have an enormous effect on the well-being of citizens and outsiders.

### 73.1 *Taxing and Spending*

An example is the issue of fairness in taxation and spending. McCaffery and Baron (2005) report several studies showing nonnormative responding (from a utilitarian perspective). People have strong and diverse views about the effects of taxation on the distribution of income. Some of our subjects favored a "flat tax" (which turned out to mean a fixed percentage of income, not a head tax). Others, to our surprise, favored so much redistribution through differential taxation (including negative taxation) that the result was close to complete equality of the resulting income.

More interestingly, perhaps, subjects generally showed isolation effects (chapter 1 by Baron in this volume). They applied their redistributive principles to whatever tax they were asked about, without considering the overall resulting distribution. For example, their judgments about the distribution of income taxes were largely unaffected by information about the existing distribution of payroll taxes, or about the distributive effects of privatization of government services.

In other studies, we found that judgments about policies were often based on their immediate effects or primary purposes, without considering side effects or long-term effects. For example, people generally favored taxes on businesses, but favored these less when they were asked who would actually pay the taxes. For further behavioral studies of tax law see chapter 13 by McCaffery in this volume.

### 73.2 *Parochialism and Nationalism*

Contrary to utilitarianism, which weighs all people equally, many people favor people in their own group, particularly co-citizens of their nation, although other groups may be favored in this way (Baron 2012a; Baron, Ritov, & Greene, in press). I use the term "parochialism" here as a technical term, but with awareness of its negative connotations. Parochialism can lead to choices that do net harm, when the benefit to the in-group is much smaller than the cost to the out-group. Examples are war and the rejection of free-trade agreements. (And the latter usually involves isolation effects as well, looking only at immediate effects on workers and ignoring both effects on consumers, and long-range effects on jobs in export sectors.)

Parochialism may be reduced by asking people about out-group members as individual people rather than as members of a group (Baron 2012a). It is also much greater for out-group harms of omission than harms of action. That is, people are much more

willing to fail to help outsiders than to fail to help insiders. Arguably this is a reasonable division of responsibility, but it seems to exist in experiments where this argument is removed.

### 7.3.3 *Naive Theories of Citizenship*

Baron (2012b; also Baron, Ritov, & Greene, in press) reports a number of studies in which people are asked explicitly about their duty as citizens. Some people think that it is their duty to vote (or otherwise act) for policies that are good for their nation, even when it makes things worse on the whole, by their own admission. For example, in one case, a voter thinks that a proposed tax on carbon fuels would be good for the world but bad for the United States (and would have no effect on the voter). Many subjects said it was their moral duty to vote against the tax. Some people go so far as to assert that it is their duty to vote on the basis of their self-interest, even when the policy they favor is worse for their nation and the world. For example, a voter thinks that a carbon tax would be bad for the world and bad for the United States but good for her, because she is in the solar energy business. Some subjects thought that this voter had a moral duty to vote for the tax. Of course both of these views are contrary to utilitarianism. People also think they would do their duty, whatever it is, so they seem willing to act in ways that they admit will make things worse, because they see this as their duty.

In the case of voting for their nation, many see this as an obligation because they think that this is why they were given the vote. In the case of voting for self, they think that voting is justified by self-interest, or that, if everyone voted this way, the majority would get what it wants (which is usually true, except that people want this to happen even when they admit it is worse on the whole). Some also think that voting for the good of the world is a betrayal of the trust bestowed on them.

People who think it is their duty to vote their self-interest are those who tend to oppose taxation and to favor government spending mostly when they see themselves as possibly benefiting from it.

I think of these views not as biases but rather as naive theories (chapter 1 by Baron in this volume). The correct theory, according to utilitarianism, is that people should vote for what is best for everyone. And, in fact, with sufficient altruistic interest in others, it is rational to do this. Despite what people think, though, it is rarely if ever rational to vote on the basis of self-interest alone.

It is difficult to find any discussion of the issue of what citizens should do in educational materials, even materials for immigrants who want to become citizens. It would be interesting to see what happens if people at least heard of the idea that voting is more rational, the more people who are affected by your vote (Baron 2012b). It is possible that such instruction is a relatively easy way to make democracy work better, in the sense of producing outcomes that truly are best for all.



## 8 OPEN ISSUES

---

I conclude with some questions for future research. Perhaps these questions have been answered, but that would be news to me.

### 8.1 Can the Law, or Anything, Improve Moral Judgment?

As noted, Robinson and Darley (e.g., 2007) and others have argued that the law should conform to intuitive judgment, lest it lose public support. Yet it is also possible that laws that do not initially conform to intuitive judgment can shape such judgment. It is difficult to disentangle cause and effect in cases when laws and public intuitions undergo change at the same time, as has happened for such issues as smoking, homosexual marriage, and recreational drugs. But it is possible that moral intuitions are responsive to changes in the law itself. If so, the concerns of Robinson and Darley may be less serious than they first appear to be. One impressive example is the effect of New York parking regulations on United Nations diplomats, who had diplomatic immunity and did not need to pay fines, yet who were responsive to changes in the law (in inverse proportion to the corruption index of their countries; Fisman & Miguel 2007). (See chapter 10 by Bilz and Nadler in this volume, for further discussion of these issues.)

A related issue is whether laws based on utilitarian principles can gain and maintain public support despite an initial conflict with moral intuition. A possible example is the use of cost-effectiveness analysis in risk regulation and healthcare. (See Baron 2008, for discussion.)

### 8.2 The Morality of Belief

Another possible point of intervention is not directly about morality but rather about belief. In times of political polarization, which sometimes results in paralysis, the two sides often differ in beliefs about natural or social sciences (evolution, economics). Often the truth is not just some compromise. Sometimes one side is completely wrong. Sometimes the wrong side can gain power. This may happen often, but a clear case is Nazi Germany. Pauer-Studer and Velleman (2011) argue that the supporters of the atrocities committed did not differ from others in moral reasoning. Many thought deeply and were troubled by what they were doing. Their support was the result of what we now see were drastically incorrect beliefs, for example, that a worldwide Jewish/Bolshevik conspiracy must be nipped in the bud and that nobody else was sufficiently concerned about it.

The existence of situations like this, which surely persist in less obvious form, raises the question of whether means of forming beliefs are a moral issue. If so, then, the sorts



of beliefs about thinking that lead to my-side bias (chapter 1 by Baron in this volume) should be included as part of the study of moral judgment.

## REFERENCES

- Adler, M. D. (2012). *Well-being and fair distribution: A framework for policy analysis*. New York: Oxford University Press.
- Baron, J. (1993a). *Morality and rational choice*. Dordrecht: Kluwer.
- Baron, J. (1993b). Heuristics and biases in equity judgments: A utilitarian approach. In B. A. Mellers and J. Baron (Eds.), *Psychological perspectives on justice: Theory and applications*, pp. 109–37. New York: Cambridge University Press.
- Baron, J. (1994). Nonconsequentialist decisions (with commentary and reply). *Behavioral and Brain Sciences*, 17, 1–42.
- Baron, J. (1996). Norm-endorsement utilitarianism and the nature of utility. *Economics and Philosophy*, 12, 165–82.
- Baron, J. (2003). Value analysis of political behavior—self-interested: moralistic :: altruistic: moral. *University of Pennsylvania Law Review*, 151, 1135–67.
- Baron, J. (2006). *Against bioethics*. Cambridge, MA: MIT Press.
- Baron, J. (2008). *Thinking and deciding*. 4th edition. New York: Cambridge University Press
- Baron, J. (2011a). Utilitarian emotions: Suggestions from introspection. *Emotion Review* (special issue on “Morality and Emotion” edited by Joshua Greene) 3, 286–87.
- Baron, J. (2011b). Where do non-utilitarian moral rules come from? In J. I. Krueger and E. T. Higgins (Eds.), *Social judgment and decision making*, pp. 261–278. New York: Psychology Press.
- Baron, J. (2012a). Parochialism as a result of cognitive biases. In R. Goodman, D. Jinks, & A. K. Woods (Eds.), *Understanding social action, promoting human rights*, pp. 203–43. Oxford: Oxford University Press.
- Baron, J. (2012b). The “culture of honor” in citizens’ concepts of their duty as voters. *Rationality and Society*, 24, 37–72.
- Baron, J., Badgio, P., & Gaskins, I. W. (1986). Cognitive style and its improvement: A normative approach. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*, vol. 3. Hillsdale, NJ: Erlbaum, 173–220.
- Baron, J., Gowda, R., & Kunreuther, H. (1993). Attitudes toward managing hazardous waste: What should be cleaned up and who should pay for it? *Risk Analysis*, 13, 183–92.
- Baron, J., Gürçay, B., Moore, A. B., & Starcke, K. (2012). Use of a Rasch model to predict response times to utilitarian moral dilemmas. *Synthese* (special issue on “Psychological Models of (Ir)rationality and Decision Making,” edited by C. Witteman & W. van der Hoek), 189, Supplement 1, 107–17.
- Baron, J., & Leshner, S. (2000). How serious are expressions of protected values? *Journal of Experimental Psychology: Applied*, 6, 183–94.
- Baron, J., & Miller, J. G. (2000). Limiting the scope of moral obligation to help: A cross-cultural investigation. *Journal of Cross-Cultural Psychology*, 31, 705–27.
- Baron, J., & Ritov, I. (1993). Intuitions about penalties and compensation in the context of tort law. *Journal of Risk and Uncertainty*, 7, 17–33.
- Baron, J., & Ritov, I. (1994). Reference points and omission bias. *Organizational Behavior and Human Decision Processes*, 59, 475–98.

- Baron, J., & Ritov, I. (2009a). Protected values and omission bias as deontological judgments. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Moral Judgment and decision making*, vol. 50 in B. H. Ross (series editor), *The Psychology of Learning and Motivation*, pp. 133–67. San Diego, CA: Academic Press.
- Baron, J., & Ritov, I. (2009b). The role of probability of detection in judgments of punishment. *Journal of Legal Analysis*, 1, 553–90.
- Baron, J., Ritov, I., & Greene, J. D. (2013). The duty to support nationalistic policies. *Journal of Behavioral Decision Making*, 26, 128–138.
- Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, 70, 1–16.
- Baron, J., & Szymanska, E. (2010). Heuristics and biases in charity. In D. Oppenheimer & C. Olivola (Eds.), *The science of giving: Experimental approaches to the study of charity*, pp. 215–36. New York: Taylor and Francis.
- Bartels, D., & Pizarro, D. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition* 121, 154–61.
- Beattie, J., & Baron, J. (1995). In-kind vs. out-of-kind penalties: preference and valuation. *Journal of Experimental Psychology: Applied*, 1, 136–51.
- Bennett, J. (1966). Whatever the consequences. *Analysis*, 26, 83–102.
- Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. New York: Cambridge University Press.
- Bucciarelli, M., Khemlani, S., & Johnson-Laird, P. N. (2008). The psychology of moral reasoning. *Judgment and Decision Making*, 3, 121–39.
- Choe, S. Y., & Min, K. (2011). Who makes utilitarian judgments? The influences of emotions on utilitarian judgments. *Judgment and Decision Making*, 6, 580–92.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Colby, A., & Kohlberg, L. (1987). *The measurement of moral judgment* (2 vols.). New York: Cambridge University Press.
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgments: Three principles of harm. *Psychological Science*, 12, 1082–89.
- Eichenberger, R., & Oberholzer-Gee, F. (1998). Rational moralists: The role of fairness in democratic economic politics. *Public Choice*, 94, 191–210.
- Feldbrugge, F. J. M. (1966). Good and bad Samaritans: A comparative survey of criminal law provisions concerning failure to rescue. *American Journal of Comparative Law*, 14, 630–57.
- Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress the spheres of justice. *Political Psychology*, 18, 255–97.
- Fisman, R., & Miguel, E. (2007). Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets. *Journal of Political Economy*, 115, 1020–48.
- Flavell, J. H. (1971). Stage-related properties of cognitive development. *Cognitive Psychology*, 2, 421–53.
- Foot, P. (1978). The problem of abortion and the doctrine of the double effect. In P. Foot, *Virtues and vices and other essays in moral philosophy*, pp. 19–32. Berkeley: University of California Press.
- Greene, J. D. (2007). The secret joke of Kant's soul, in W. Sinnott-Armstrong (Ed.), *Moral psychology*, vol. 3: *The neuroscience of morality: Emotion, disease, and development*, pp. 36–79. Cambridge, MA: MIT Press.

- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*, 364–71.
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*, 1144–54.
- Haidt, J. (2001). The emotional dog and its rational tale. *Psychological Review*, *108*, 814–34.
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, Fall, 55–66.
- Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or, Is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*, 613–28.
- Hare, R. M. (1981). *Moral thinking: Its levels, method and point*. Oxford: Oxford University Press (Clarendon Press).
- Harris, R. J., & Joyce, M. A. (1980). What's fair? It depends on how you phrase the question. *Journal of Personality and Social Psychology*, *38*, 165–79.
- Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience*, *7*, 392–402.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986a). Fairness and the assumptions of economics. *Journal of Business*, *59*, S285–300.
- Knobe, J. & Nichols, S. (2008). *Experimental philosophy*. New York: Oxford University Press.
- Knudsen, T. (2003). Simon's selection theory: Why docility evolves to breed successful altruism. *Journal of Economic Psychology*, *24*, 229–44.
- Kogut, T., & Ritov, I. (2005). The “identified victim” effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, *18*, 157–67.
- Kohlberg, L. (1963). The development of children's orientations toward a moral order. I. Sequence in the development of human thought. *Vita Humana*, *6*, 11–33.
- Kuhse, H. (1987). *The sanctity of life doctrine in medicine: A critique*. Oxford: Oxford University Press.
- Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin*, *85*, 1030–51.
- Lerner, M. J., & Simmons, C. H. (1966). Observer's reaction to the “innocent victim”: Compassion or rejection. *Journal of Personality and Social Psychology*, *4*, 203–10.
- McCaffery, E. J., & Baron, J. (2005). The political psychology of redistribution. *UCLA Law Review*, *52*, 1745–92.
- McNaughton, D., & Rawling, P. (1991). Agent relativity and the doing-happening distinction. *Philosophical Studies*, *63*, 167–85.
- Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. New York: Cambridge University Press.
- Moretto, G., Ládavas, E., Mattioli, F., & di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *22*, 1888–99.
- Nabi, R. L. (2002). The theoretical versus the lay meaning of disgust: Implications for emotion research. *Cognition & Emotion*, *16*, 695–703.
- Pauer-Studer, H., & Velleman, J. D. (2011). Distortions of normativity. *Ethical Theory and Moral Practice*, *14*, 329–56.
- Paxton, J. M., Ungar, L., & Greene, J. D. (2011). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*, 163–77.
- Piaget, J. (1932). *The moral judgment of the child*. Glencoe, IL: Free Press.

- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
- Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making*, 3, 263–77.
- Ritov, I., & Baron, J. (1999). Protected values and omission bias. *Organizational Behavior and Human Decision Processes*, 79, 79–94.
- Robinson, P. H., & Darley, J. M. (2007). Intuitions of justice: Implications for criminal law and justice policy. *Southern California Law Review*, 81, 1–67.
- Roth, A. E. (2007). Repugnance as a constraint on markets. *Journal of Economic Perspectives*, 21, 37–58.
- Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, 15, 165–84.
- Royzman, E. B., Leeman, R. F., & Baron, J. (2009). Unsentimental ethics: Toward a content-specific account of the moral-conventional distinction. *Cognition*, 112, 159–274.
- Rudski, J. M., Osei, W., Jacobson, A. R., & Lynch, C. R. (2011). Would you rather be injured by lightning or a downed power line? Preference for natural hazards. *Judgment and Decision Making*, 6, 314–22.
- Schleifer, M., Shultz, T. R., & Lefebvre-Pinard, M. (1983). Children's judgements of causality, responsibility and punishment in cases of harm due to omission. *British Journal of Developmental Psychology*, 1, 87–97.
- Shavell, S. (2004) *Foundations of economic analysis of law*. Cambridge, MA: Belknap Press of Harvard University Press.
- Singer, P. (2005). Ethics and intuitions. *Journal of Ethics* 9, 331–52.
- Slovic, P. (2007). "If I look at the mass I will never act": Psychic numbing and genocide. *Judgment and Decision Making*, 2, 79–95.
- Small, D. A., & Loewenstein, G. (2005). The devil you know: The effects of identifiability on punishment. *Journal of Behavioral Decision Making*, 18, 311–18.
- Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, 27, 76–105.
- Sunstein, C. R., Schkade, D., & Kahneman, D. (2000). Do people want optimal deterrence? *Journal of Legal Studies*, 29, 237–53.
- Suter, R. S., & Hertwig, R. (2011). Time and moral judgment, *Cognition*, 119, 454–58.
- Tetlock, P. E., Lerner, J., & Peterson, R. (1996). Revising the value pluralism model: Incorporating social content and context postulates. In C. Seligman, J. Olson, & M. Zanna (Eds.), *The psychology of values: The Ontario symposium*, vol. 8, pp. 25–51. Hillsdale, NJ: Erlbaum.
- Thaler, R. H. (1988). The ultimatum game. *Journal of Economic Perspectives*, 2, 195–206.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. New York: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–58.
- Ubel, P. A., Baron, J., and Asch, D. A. (2001). Preference for equity as a framing effect. *Medical Decision Making*, 21, 180–89.
- Ubel, P. A., and Loewenstein, G. (1996). Distributing scarce livers: The moral reasoning of the general public. *Social Science and Medicine*, 42, 1049–55.
- Walster, E., Walster, G. W., & Berscheid, E. (1978). *Equity: Theory and Research*. Boston: Allyn & Bacon.
- Zamir, E., & Ritov, I. (2012). Loss aversion, omission bias, and the burden of proof in civil litigation. *Journal of Legal Studies*, 41, 165–207.



PART II

---

**BEHAVIORAL  
ECONOMICS AND  
THE LAW**

*An Overview and Critique*

---





## CHAPTER 4

---

# THE IMPORTANCE OF BEHAVIORAL LAW

---

THOMAS S. ULEN

## 1 INTRODUCTION

---

LAW is a behavioral system.<sup>1</sup> It seeks to shape human behavior—to regulate, to incentivize, to nudge people to behave in some ways and not to behave in others. In recent decades, legal scholars have come to understand that human behavior can be deliberately shaped through many mechanisms beyond the administration of technically managed formal rules—through mechanisms including economic incentives, social norms, and psychological framing. All of these tools provide opportunities for manipulating behavior, and often they can be used as substitutes, both for each other and for familiar legal mechanisms like property rights, criminal sanctions, or civil liability.

Moreover, legal scholars have learned that the leading theory of human behavior—rational choice theory (RCT), which had been developed by economists in the 1950s and had come to the law through the rise of law and economics—has significant shortcomings. A body of literature, originally from social and cognitive psychology and increasingly from legal scholars, has relied on laboratory experiments, field experiments, and analysis of archival data to demonstrate that human decision-making does not typically result in behaviors predicted by RCT. Human beings make mistakes in their judgment and decision-making. These mistakes are predictable, pervasive, and difficult to correct.

The central project of behavioral law is to undertake legal analysis on the basis of the findings of these behavioral studies.<sup>2</sup>

<sup>1</sup> Much of this section was written by my friend and colleague, Arden Rowell.

<sup>2</sup> We have chosen to characterize the subject matter about which we are writing as “behavioral law” or “behavioral legal studies.” The more common term might be “behavioral law and economics,” but we are not comfortable with that term. Why? While it is true that all of this began with law

What makes behavioral law distinctive as a discipline, if—as we have learned—tools from other behavioral disciplines can so frequently substitute for familiar legal mechanisms? In our view, the study of *law* and behavior is distinctive because law seeks to intentionally shape people’s behavior by reference to some measure of *social* good. This distinguishes it from fields based on manipulation for individual ends, such as advertising or marketing; from fields oriented on thought rather than behavior, such as much of philosophy; and, importantly, from other behavioral fields that are based on descriptive or evaluative treatments of behavior, such as anthropology, sociology, psychology, and economics, which do not necessarily seek to control what they study nor to engage in normative policy analysis.

This intentional relationship with respect to social good is what we see as distinctive about the study of law and human behavior. The complexity of the relationships between law and behavior is what makes collaborative projects like this *Handbook* so valuable. No one alone can hope to master all of the literatures and modes of thought that might inform the purposeful regulation of human behavior. In studying law and behavior, then, we all must rely upon the thoughts and work of colleagues who know more than we do about something important. The downside of this is the pervasive and inescapable feeling that one can never be a master of one’s own discipline. But the upside is a sense of community and camaraderie, and—we think—a helpful (if occasionally painful) pressure towards intellectual humility, which helps behavioral law scholars as they edge incrementally closer to accurate descriptions of the forces shaping human behavior and to understanding how those forces can be deliberately harnessed.

What can law and behavior scholars hope to do when evaluating work across the range of disciplines—from economics to anthropology to philosophy to sociology to history to psychology? We think the most important skills within law and behavior are the abilities to identify, evaluate, and criticize the alignment of methods and goals. To that end, the remainder of this chapter is given over to three parts, each of which addresses one aspect of these inquiries.

In section 2, we provide a brief overview of the existing literature in behavioral law. Other chapters and authors will have contextualized law and behavior as a field and will have introduced and criticized several of the more important methods that have been developed to address the challenges of synthesis posed by legal behavioral studies. So, our focus in section 2 can be on what distinguishes behavioral law from what has come before and on its importance for discussing legal issues.

and economics in the late 1970s, the next forty years have seen dramatic growth in the field. It is difficult (and perhaps foolhardy) to summarize this growth, but we think that it centrally consists of incorporating techniques and methods and findings from a wide variety of social sciences into the study of law. Economics might once have had pride of place among the disciplines contributing, but we think that the cooperative and collaborative nature of the many social sciences now contributing to the study of law is the distinctive characteristic today. The University of Illinois College of Law recognized this development by changing the name of its Law and Economics Program to the Program in Law, Behavioral, and Social Sciences.

In section 3, we consider several particular areas of the application of behavioral legal insights in order to develop a typology of uses of behavioral studies in law. We use this section to outline four different ways that we see behavioral studies used in law and behavior: to inform predictive and descriptive models, to persuade people to change their behavior, to evaluate decision-making quality, and to prescribe changes in policy that will lead to changes in behavior. We see each of these uses as creating different methodological pressure, and argue that law and behavior scholars add significant value when they seek to ensure that the methodologies underlying studies align with the uses to which those studies are put.

In section 4, we address ongoing challenges to the field of law and behavior, and do our best to sketch where we see these challenges as fundamental to the discipline, and where we see potential retorts. We see constructive engagement with these criticisms as critical to maintaining the rigor and relevance of behavior law as a field of inquiry.

We conclude with a few reflections about the future of law, behavior, and social science, gesturing at important emerging questions, remaining limitations, and new methodologies.

## 2 THREE EXAMPLES OF BEHAVIORAL FINDINGS

---

Examples of the findings of behavioral experiments, both laboratory and field, are very well known within the social science community, including the legal academy. In this section I shall give only three examples with the intention of using them in the next section to show the importance of the behavioral literature for the analysis of law.<sup>3</sup>

There are three central points to stress about these examples and about the behavioral literature generally. First, these examples are not theoretical; they are empirical. That is, they do not begin from a hypothesis about how human beings might make decisions in the circumstances or about the matters under investigation. Rather, these experiments all focus on the extent to which RCT is a good explanatory model of how people actually behave.

Second, in these examples and in almost every other example that one could cite from the behavioral literature, the experimental results fail to confirm the predictions of RCT. Just as importantly, the experimental findings do not show that human behavior is chaotic or unpredictable. Quite to the contrary, the behavioral experiments find that human choice behavior is predictable: Most humans behave in a similar ways in similar circumstances; those ways, however, are not those predicted by RCT. It follows that using RCT to predict some behavior may lead to mispredictions.

<sup>3</sup> Some of this section appeared in a slightly different form in a paper presented at a Liberty Fund Conference in February, 2012.

Third, because behavioral experiments suggest that human beings do not behave as RCT predicts, people can be said to make “mistakes” in their choices. I resist calling these mistakes “irrationalities”; they do, nonetheless, suggest that people do not typically make decisions that enhance their welfare as much as they might. This fact creates a need to figure out how, if at all, one might allow people to avoid these mistakes so as to enjoy greater well-being.

## 2.1 Framing

In a famous experiment, Amos Tversky and Daniel Kahneman (1981) asked experimental subjects to choose between various prevention strategies for dealing with a very serious disease that might affect a patient population of 600 people. The first choice was between these two options:

*Option A:* This program will save 200 lives.

*Option B:* This alternative program has a 33% chance of saving all 600 subjects and a 67% chance of saving no one.

Both programs have an expectation of saving 200 lives. But because Option A's saving is a certainty and Option B's is probabilistic, one would expect everyone who is risk-averse to prefer A to B. When Tversky and Kahneman presented these options to various subjects, approximately 70% preferred A to B and approximately 30% preferred B to A.

Then, Tversky and Kahneman presented another choice on similar facts to different groups. The characteristics of these various groups were indistinguishable from those of the groups who chose between Options A and B. This second choice was between these two options:

*Option C:* This program will result in the deaths of 400 people.

*Option D:* This alternative program has a 33% chance of no one's dying and a 67% chance that all 600 people will die.

Again, both programs have the same expectation as Options A and B—saving 200 lives—but here the choices focus not on lives *saved* but on lives *lost*. Notice, also, that Options C and A amount to the same thing, as do Options D and B. For all those reasons, the investigators' expectation was that groups put to the choice between C and D would behave identically to those put to the choice between A and B. When, however, Tversky and Kahneman presented the choice between C and D, they found that 78% preferred D to C and only 22% preferred C to D.

Kahneman and Tversky called this odd result the “framing effect.” This effect is closely connected to “prospect theory,” which gives a comprehensive explanation for the choice phenomena the public health example illustrates (Kahneman and Tversky 1979;

see also Zamir 2012; Barberis 2013). There are several important implications of this effect. One theoretical implication is that it calls into question an implicit assumption in RCT that decision-makers' choices are invariant to the manner in which information is presented to them. Another implication is the very practical one that those choices can depend crucially on how the information is "framed." That is a vital point for policy-makers to bear in mind. Additionally, it is a point that some framers might use to their own advantage by, for example, structuring choices to get an outcome that they prefer while giving the decision-makers the illusion that they have chosen freely (Riker 1986).

## 2.2 Default Rules

A default rule is a starting rule that will be in effect unless the party or parties facing the rule change it. If the transaction costs facing the chooser are identical and low regardless of the default rule, then one might anticipate that whatever the default is will not keep the chooser from reaching his or her most desired outcome. That is, the choice of default rule, like the choice of an initial assignment of entitlement, would seem to obey the Coase Theorem: An efficient outcome will result, regardless of the starting point, if transaction costs are low (Cooter and Ulen 2012).

In an important article, Cass Sunstein and Richard Thaler (2003, 1064) give an example of how one might think about default rules in light of behavioral economics:

[C]onsider the cafeteria at some organization. The cafeteria must make a multitude of decisions, including which foods to serve, which ingredients to use, and in what order to arrange the choices. Suppose that the director of the cafeteria notices that customers have a tendency to choose more of the items that are presented earlier in the line. How should the director decide in what order to present the items? To simplify, consider some alternative strategies that the director might adopt in deciding which items to place early in the line:

1. She could make choices that she thinks would make the customers best off, all things considered.
2. She could make choices at random.
3. She could choose those items that she thinks would make the customers as obese as possible.
4. She could give customers what she thinks they would choose on their own.

Sunstein and Thaler dismiss Options 2 and 3. They identify Option 1 as paternalistic and Option 4 as "what many anti-paternalists would favor." But they are skeptical that most consumers have well-formed preferences about how they would like the food arranged. They suggest that if most consumers would choose differently if the food were to be arranged in different orders, then one may conclude that preferences are not exogenous but are endogenous in some manner that seems beyond the subjects' control and knowing.

It is possible, they argue, that the goal of the cafeteria director is to maximize profits. If he or she is in a marketplace in which there are attractive alternative

eating venues outside the organization, then competition imposes a constraint on profit-maximization. But it is also possible that the cafeteria has a degree of market power over its customers: Perhaps it is a school from which the students are forbidden to leave campus or a dormitory with which the residents have a prepaid eating contract.

Sunstein and Thaler's central point in the cafeteria example is to suggest that paternalism (Option 1) in the arrangement of the cafeteria's choices is a mere description, not a pejorative. There is no coercion, they suggest, in how the food is ordered:

Would anyone object to putting the fruit and salad before the desserts at an elementary school cafeteria if the result were to increase the consumption ratio of apples to Twinkies? Is this question fundamentally different if the customers are adults? (p. 1166)

And

Once it is understood that some organizational decisions are inevitable, that a form of paternalism cannot be avoided, and that the alternatives to paternalism (such as choosing options to make people worse off) are unattractive, we can abandon the less interesting question of whether to be paternalistic or not, and turn to the more constructive question of how to choose among the possible choice-influencing options. To this end we make two general suggestions. First, programs should be designed using a type of welfare analysis, one in which a serious attempt is made to measure the costs and benefits of outcomes (rather than relying on estimates of willingness to pay). Choosers should be given more choices if the welfare benefits exceed the welfare costs. Second, some results from the psychology of decisionmaking should be used to provide *ex ante* guidelines to support reasonable judgments about when consumers and workers will gain most by increasing options. (p. 1166)

Sunstein and Thaler are well aware of the danger in letting someone other than the end-consumers make these decisions. To minimize that danger, they suggest that if those in authority are to make decisions about, say, how the food is to be ordered in the organization's cafeteria, they do so in a transparent manner that the end-consumers can reject.

Perhaps an even better example of Sunstein and Thaler's point about paternalistic default rules comes from a fascinating empirical study of default rules for organ donation.

The science and, therefore, availability of organ transplantation, first practiced with corneas in 1905, is one of the most significant medical advances of the last fifty years. To take one example, Dr. Christiaan Barnard performed the first heart transplant in 1967. In 2008 there were slightly more than 2,000 heart transplants in the United States alone (Johnson et al. 2010).

The central economic issue in organ transplantation is the huge excess demand for transplantable organs. There are approximately 100,000 people on the various waiting lists for organ transplants in recent years and only 30,000 organs available for transplant. Roughly 6,000 people on the waiting list die each year without having received a transplant. At the same time that the number of transplants has been

increasing—doubling between 1988 and 2006—the number of people on the waiting list has been growing much more rapidly—increasing approximately sixfold between 1988 and 2006.

How should we deal with this persistent and widening excess demand? US federal law (the National Organ Transplantation Act of 1984, Pub. L. 98-507, amended in 1988 and 1990) prohibits an explicit market payment for an organ; so, the most obvious economic proposal for increasing supply is off the table. Some commentators have explored the possibility of avoiding this proscription by giving in-kind gifts, such as monetary donation to a legitimate charitable organization, in exchange for donating one's organs for transplantation (Orentlicher 2009).

A possible method of increasing the supply of transplantable organs is to change the default rule for organ donation. There are two principal defaults—one in which individuals are presumed to consent to have their organs harvested for transplantation upon their death, and one in which individuals are presumed *not* to consent to having their organs harvested. Of course, these are merely defaults: Individuals may choose to move away from them. For instance, in the presumed consent regime someone who does not want to donate her organs can simply opt out by, for example, signing the back of her driver's license on the appropriate line that indicates that she is not an organ donor. Similarly, in the regime in which the default is no donation, an individual can opt in by, for example, signing the back of his driver's license to indicate that some or all of his organs are available for transplantation. If the costs of contracting away from either of these defaults—either opting in or opting out—are minor (as it, in fact, is), then one might predict that the ultimate number of organs donated for transplantation would be roughly the same under either default. That is, the choice of default might have no effect on the supply of transplantable organs.

But there is persuasive empirical evidence to suggest that the choice of default *does* affect the number of donated organs (Johnson and Goldstein 2003). Here are the facts: 85% of Americans approve of organ donation and express an intention to donate. However, fewer than 50% have actually made a decision to donate, and only 28% have signed a donor card or in some other way made their intention explicit. Surveys in Germany, Spain, and Sweden have found that individuals in those countries have exactly the same feelings about donation and have done just as little about it as have those in the United States.

To see if they could explain these facts, Johnson and Goldstein conducted three experiments. In the first, they used an online survey with 161 respondents. Those participants were told that they had recently moved into a new state and had to adopt a policy with respect to their organ donations. One-third of the respondents was told that the prevailing rule was *not* to be a donor but that they could opt in to become a donor. Another third was told that the prevailing rule was to be an organ donor but that they could opt out of that rule. A final third of the subjects was told that there was no default rule; the subject had to express a preference by opting into becoming a donor or opting out of being a donor.



The results were revealing. Subjects' donation rates were about twice as high when they had to opt out as when they had to opt in, even though the cost of expressing either option was the same.<sup>4</sup> In the neutrality condition, 79% chose to donate, only slightly less than the percentage of respondents who chose to donate when there was presumed consent with opt out.<sup>5</sup>

These results would suggest that, all other things equal, and following the logic of libertarian paternalism articulated by Sunstein and Thaler, paternalism would argue for a policy of presumed consent with opt out, either because that conforms to individuals' "true" preferences or that having a greater supply of donated organs is socially desirable.

Johnson and Goldstein also compared organ donation rates across European countries that differ according to whether they have presumed consent with an opt-out rule or no consent with an opt-in rule. The four countries that have an opt-in rule—Denmark, the Netherlands, the United Kingdom, and Germany—have much lower donation rates, ranging from 4.25% in Denmark, to 12% in Germany, to 17.17% in the UK, to 27.5% in the Netherlands.

The six countries that have opt-out rules—Austria, Belgium, France, Hungary, Poland, Portugal, and Sweden—have much higher rates of donation, ranging from a low of 85.9% to a high of 99+%.

It does not appear to be the case that religion, total population, ethnic diversity, or income per capita are statistically significant explanatory variables for these stark differences in actual donation rates.

Finally, the authors also ran a regression in which the dependent variable was the actual number of donations (presumably scaled by population) and the independent variables included a dummy variable for whether the prevailing rule on donation was one of opt in or opt out. Their results were strong: "When donation is the default, there is a 16.3% increase in donation."

Interestingly, beginning in the 1960s about two-thirds of the states adopted presumed-consent laws for some body-part donations, such as corneas, pituitary glands, and some tissues and organs. The 1987 Uniform Anatomical Gift Act adopted presumed consent for a limited number of body-part donations. However, the 2006 Revised UAGA eliminated the presumed consent provision of the 1987 act and substituted a requirement of actual consent by the donor or immediate family members. About thirty-three states and the District of Columbia have passed the Revised UAGA but a few of them have retained presumed consent for corneas only (Orentlicher 2009).

<sup>4</sup> The donation rate was 42% when respondents had to opt in and 82% when they had to opt out of presumed consent.

<sup>5</sup> There is still a bit of a puzzle about why these rates are so dramatically different when the costs of opting in or opting out are equal. Johnson and Goldstein propose three possibilities: (1) respondents construe the default as a recommendation from policymakers; (2) accepting a default is virtually costless, while opting in or out is costly, and people seek to minimize their costs; and (3) the default is the status quo and people have a bias toward the status quo. Neither of the last two possibilities seems to explain the results of the neutrality version of the experiment.

RCT argues that in some—perhaps, many—circumstances (especially those of low transaction costs) default rules do not matter to the efficient allocation and use of resources. In contrast, behavioral experiments demonstrate that default rules matter to choice: For whatever reason, most people follow the default rule in a given choice. As a result, where policymakers and others set the default may matter significantly to the choices that those facing the default make. As was the case with framing, people may feel the illusion of choosing freely when, in fact, their choice has been partly determined by the default rule (chapter 11 by Zamir and chapter 12 by Korobkin in this volume).

### 2.3 Affective Forecasting

The recent literature on the psychology and economics of happiness has generated some remarkably interesting results that are, at several points, related to the behavioral economics literature (Diener and Seligman 2004; Bronsteen, Buccafusco, and Masur 2010). One of those points of intersection has to do with “affective forecasting,” the human ability to predict future emotional states—that is, the pleasure or pain that some as-yet-unexperienced event or good will give us. The thrust of this literature is that we are very poor at these predictions (Gilbert 2008). For example, suppose that you love to read; it is your principal source of pleasure when you are not working. If you were asked to predict how you would feel if you lost your eyesight, you would no doubt predict that it would very significantly reduce your subjective well-being. However, if you were, in fact, to lose your eyesight, your prediction of a large negative impact on your subjective well-being would probably not be correct. There might well be a reduction in well-being after the initial loss of eyesight, but that reduction is not likely to be permanent. If most studies are to be believed, you will return to your preloss level of subjective well-being within a year and perhaps sooner (Brickman et al. 1978).<sup>6</sup>

Why are we so bad at forecasting the things that will give us pleasure and pain? We underestimate our ability to adapt to both good and bad events, which ability is very strong. Indeed, our ability to adapt is so strong that some psychologists have hypothesized that each of us has a “set point” of happiness to which we eventually return whenever circumstances push us away from that setting—much as a thermostat causes the ambient temperature to return to a preset temperature whenever events cause the ambient temperature to deviate from that setting. No one is certain yet how that set point of individual happiness comes about—what, for instance, the roles of heredity, environment, early life circumstances, and the like might be. It is, however, fairly

<sup>6</sup> There have been many confirming studies since 1978 (see Gilbert 2006). In most instances of both adverse and beneficent events, the loss or increase in subjective well-being has disappeared within twelve months, causing the level of well-being to return to its level before the (bad or good) event. But there have also been other studies that show that for some losses, such as those of a spouse or partner, a child, and a job, the loss in subjective well-being is large and the recovery is incomplete.

certain that it is difficult to reset one's happiness level to a permanently higher level. Doing so involves somehow recognizing the adaptive process and circumventing it (Seligman 2011; Dunn, Gilbert, and Watson 2011; Dunn and Norton 2013). If we do not take those steps to move our set point higher, then the only way in which we are likely to be able to have greater subjective well-being is to have ever more expensive and expansive experiences and goods, to get on what psychologists call the "hedonic treadmill," in which to be happier, you have to keep going faster and farther (Schkade and Kahneman 1998; Kahneman et al. 2006).

There are important legal and policy implications of our difficulties with predicting the pleasure and pain that future events will give us. I shall consider two of those implications in section 3.2 below, those arising in criminal law and tort law. But there are other areas of law and policy in which considerations of our difficulties with affective forecasting are important. For example, in the United States we allow victim impact statements (testimony by friends and relatives of a murder victim, say, about how the crime has and will affect their lives and emotions) to be made at a sentencing hearing in criminal prosecutions to determinate the degree of wrong perpetrated by the defendant. These statements have been found to be very influential in capital sentencing, even though the literature on predicting future emotional states cautions against putting a great deal of credence in what the victims say (Blumenthal 2005). There are also implications for the legal treatment of euthanasia and physician-assisted suicide and advance directives about future medical interventions (Blumenthal 2005).

### 3 BEHAVIORAL LAW AND ECONOMICS

---

In this section I explore the implications of some findings from the behavioral literature for law. There has already been much writing on this topic. For example, Richard Thaler and Cass Sunstein have written persuasively on the role that gentle "nudges"—as in the setting of default rules, as in the Save More Tomorrow program (Thaler and Benartzi 2004)<sup>7</sup>—can play in leading people to make welfare-enhancing choices.

A central purpose of the examples developed here is to suggest that while nudges are useful in some selected situations (as in increasing organ donations), they are not useful in many other situations. There are, that is, situations in which the deviations from the predictions of RCT discovered in the behavioral studies call for a much more intrusive intervention in private decision-making and a thorough rethinking of some core aspects of law (see also Bubb and Pildes 2014). For instance, there may be a much stronger case to be made in favor of hard shoves or even paternalism than economists

<sup>7</sup> "Save More Tomorrow" is a program in which employees agree to "allocate a portion of future salary increases toward retirement savings." <http://www.chicagobooth.edu/capideas/summer02/savemoretomorrow.html>.

have heretofore been comfortable with (Zamir 1998; Zamir and Medina 2010, ch. 10; Conly 2013; and Willis 2014).

I consider two examples—the possibility of financial literacy education and the implications of difficulties in affective forecasting for minimizing the social costs of accidents and for deterring criminal activity.

### 3.1 Financial Literacy and the Limitations of Information Dissemination

The standard rational choice model of human behavior predicts that rational human beings will save and consume income at each point in time so as, roughly, to smooth out their consumption pattern over their lifetime. So, for example, if the typical pattern of one's lifetime earnings can be expected to be low in one's early years of employment, to rise to a peak in, say, one's fifties, then to decline as one approaches retirement, and to go to zero when one retires, then a rational person will—again, roughly—borrow as a younger person so as to be able to spend more than one's income (by, for example, taking out a mortgage on a house), save at some point in his middle age, and then dis-save (by drawing down the savings so as to maintain the same preretirement level of consumption) when one's working income stops.

Because rational people foresee most of these variations, they may not need significant help in planning for them. Since the mid-1930s, however, the US federal government (through the Social Security Administration) has required almost 95% of all working Americans to contribute to a plan to provide for their retirement income (and to provide other benefits in the event of disability or death). In addition, the US government gives modest tax advantages to private individual savings for retirement (in a traditional Individual Retirement Account or a Roth IRA or the various supplemental plans—such as 401(k) and 403(b)(7) plans—available through private and public employers).<sup>8</sup> Despite the compulsory nature of the Social Security System and the generous benefits of the associated Medicare system (and the future funding problems of those programs), the general political stance toward them seems to be that they are safety nets meant to catch and support those who need modest help in planning against unforeseen contingencies. Most people are presumed to take their own steps toward their retirement and medical care.

<sup>8</sup> These various plans for retirement planning are explained extensively at nearly every website for financial planning. See, for example, <http://www.aarp.org/work/retirement-planning/> or [https://www.fidelity.com/retirement-planning/overview?imm\\_pid=1&immid=00702&imm\\_eid=e40582582&buf=999999&imm\\_pid=1&immid=00642&imm\\_eid=e12589691&buf=999999](https://www.fidelity.com/retirement-planning/overview?imm_pid=1&immid=00702&imm_eid=e40582582&buf=999999&imm_pid=1&immid=00642&imm_eid=e12589691&buf=999999). The central point to note here is that there is a great deal of information available to individual decision-makers about their choices for saving for retirement.

One of the most solid findings of the behavioral research is that people do not do a very good job of dealing with time in their consumption and saving decisions. They engage in hyperbolic discounting and myopic behavior, placing too much emphasis on current pleasure and undervaluing future costs. For example, the average American, who will retire at sixty-two and live to be eighty, has accumulated only about \$45,000 in savings for her retirement by the time she stops working (Schwartz Center for Economic Policy Analysis 2012).<sup>9</sup> More than one-third of those over sixty-five in the United States retire on Social Security for their entire retirement income; more than half of retired married couples and nearly three-quarters of unmarried retirees in the United States get at least half their retirement income from Social Security (Social Security Administration 2013).<sup>10</sup>

There is other evidence, beyond the merely financial, that people do not do well in making current decisions that have future costs and benefits. For instance, almost one-third of all Americans today are obese, even though the evidence is that this extra weight can have dire health consequences later in life (Centers for Disease Control and Prevention 2012). Most people undervalue the future benefits of current exercise. They underestimate the future health costs of the current consumption of high-cholesterol foods, like red meat and cheese; of cigarettes; and excessive alcohol. They exaggerate the current pleasure and understate the future pain of taking addictive drugs, including not just illegal drugs like cocaine and heroin but legal drugs like Oxycontin and Adderall.

Are there nudges that we can apply to help with these problems? Or do these problems require more serious shoves? One possibility is that people will do better with financial planning if they are given training to make them financially literate. This would involve teaching people about such things as compound interest and future values, discounting to present value so as to be able to compare current and future dollar amounts, annuities, the elements of financial risk, and the characteristics of various financial instruments (such as individual stocks, bonds, mutual funds, and the like).

There is certainly a need for this teaching. Annamaria Lusardi and Olivia Mitchell have studied financial knowledge for years (Lusardi and Mitchell 2009). They recently reported that a financial literacy test was administered to US adults 50 years and older. The adults were asked to answer the following three questions:

<sup>9</sup> The more detailed figures are even more startling. For the two lowest personal income quartiles (those with annual incomes of less than \$27,500), the mean total savings for those who are fifty to sixty-four years of age in the lowest quartile are \$16,034, with a median of \$0; for those who are fifty to sixty-four in the next highest quartile, the median total savings are \$21,606, with a median of \$0. For those in the highest quartile (annual personal income of greater than \$52,201) the mean total savings are \$105,012, with a median of \$52,000. For the purposes of rough accounting, most financial planners suggest total savings equal to twenty times one's annual income on the eve of retirement as a prudent amount.

<sup>10</sup> It is worth noting that those who are eligible for Social Security in the United States also receive health benefits from Medicare, the federal-government-sponsored provider of hospitalization, routine physician care, and drug services. There is a modest and progressive charge for those services, but the value of those health benefits is considerably larger.

1. Suppose you had \$100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow: more than \$102, exactly \$102, less than \$102?
2. Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy more than, exactly the same as, or less than today with the money in this account?
3. Do you think that the following statement is true or false? "Buying a single company stock usually provides a safer return than a stock mutual fund."

Only about half of those given these questions could answer two of the three correctly.

Clearly, it would be better if people knew more about these basic matters of finance. They would consume and save more wisely; they might be less subject to financial chicanery; there might be fewer individual bankruptcies; and there might be less contention within families about financial matters (one of the principal sources of familial breakup). In addition, there is a strong positive correlation between financial literacy and personal wealth, and the relationship is probably causal, not merely correlational. Indeed, there is some recent evidence that variations in the degree of financial literacy is an important factor in explaining the large degree of income and wealth inequalities in the United States (Lusardi, Michaud, and Mitchell 2012).

Some states and the federal government have recognized the social value of increased financial literacy and have instituted programs directed at those who are thought to need to have that increased literacy. For example, the 2005 Bankruptcy Abuse Prevention and Consumer Protection Act (BAPCPA) "mandated that individuals seeking bankruptcy protection must undergo a credit counseling course prior to filing for bankruptcy as well as a financial education course before receiving a discharge of their debts" (Jimenez et al. 2013, 453).<sup>11</sup> The state of Illinois has encouraged school districts to institute financial literacy courses and has sponsored an online course available at the Department of Financial and Professional Regulation website.

The feeling that more information will help people make better decisions is widespread, natural to academics, and easy for broad political coalitions to support. And yet there is some evidence that information dissemination is not a very effective general regulatory strategy and, with respect to financial literacy, more expensive than beneficial and not very effective (Fernandes et al. 2013).

Omri Ben-Shahar and Carl Schneider (2011) have written a strongly persuasive critique of information dissemination as a general regulatory strategy. Mandated disclosure plans are expensive to comply with; consumers, who are presumed to benefit from the information disclosures, find themselves overwhelmed by the amount of

<sup>11</sup> This article proposes a randomized controlled trial that will shed some light on whether these mandatory financial literacy courses have had a positive effect. The evidence on that effect has, to date, been ambiguous because there is no control group under BAPCPA: Everyone seeking bankruptcy relief and discharge must take the courses.

information with which they must deal; consumers have a limited ability to retain the information in working memory (typically retaining no more than a third of information disclosed to them); and the mandatory information can have undesirable unintended consequences—for instance, crowding out useful information, harming competition, and fostering inequity.

With respect to financial literacy, Lauren Willis argues that attempts to foster wider knowledge of financial matters have not been effective (Willis 2011). She argues that courses in financial literacy have failed because they are expensive, and financial services providers change their products frequently, making it necessary to revise courses frequently.

On the other hand, some scholars have found some modest benefits of financial literacy courses. One group recently reported that employer-delivered financial literacy information has generally been found to help employees make better financial decisions. In part this is because the information is tailored to the needs of the particular employees and in part because the information is typically delivered at the particular time that a decision is being made (Clark, Morrill, and Allen 2012). Another group studied Junior Achievement<sup>12</sup> members in Southern California who had been randomly divided into those who received financial literacy training and those who did not. Both groups then went through a financial simulation. The savings rates of students who went through the financial course twice were “four times higher after the education, they paid off their debt faster, and they spent considerably less on entertainment and dining out” (Carlin and Robinson 2012, 306).

One unusual means of informing people about financial matters is video games (*The Economist* 2013). The Doorways to Dreams (D2D) Fund, a Boston-based nonprofit organization that seeks to increase the financial literacy of poor and middle-class members, introduced a computer game in 2009 called *Bite Club*. The player of the game manages a bar for vampires and must fulfill orders for blood types for the customers while putting away enough money for retirement—on the theory that vampires are “undead” and will live forever. D2D recently reported a randomized control trial in which, first, all participants took a financial knowledge test. Then, half the participants played a game like *Bite Club*, and the other half read an instructional pamphlet. Finally, both groups took a second test on financial knowledge. Both groups did better on the second test, with those who had read the pamphlet doing slightly better. D2D stressed that those who played the video game apparently learned almost as much as did those that read the pamphlet. Perhaps D2D is onto a more effective way to disseminate information.

The upshot of these various studies is that there are no simple nudges that seem to work to improve financial decision-making. Whether there are more intrusive regulatory devices or more clever methods, based on something like *Bite Club*, I do not know.

<sup>12</sup> Junior Achievement is a nonprofit organization in the United States and elsewhere founded in 1919 for teenaged boys and girls, who work with local businesspeople and organizations to learn, through realistic situations, about financial matters, business practices, and entrepreneurship.



It could be that the new Consumer Finance Protection Bureau (CFPB) will introduce some effective and novel means of helping consumers deal with financial matters (Levitin 2013).

### 3.2 Affective Forecasting in Tort Liability and Criminal Law

Subsection 2.3 above described the difficulties that human beings have with affective forecasting (predicting how future events will impact their subjective well-being). One implication of this finding is that left to their own devices, most people will make life plans and undertake actions and consume things that will not ultimately make them better off. And they may incur costs to avoid losses that they think will be catastrophic to their well-being but that would, in the event, prove not to be so devastating.

A related implication is that if people are prone to these errors and if, as seems to be the case, they do not necessarily get better at affective forecasting over time or through experience, then there may be a case for paternalistic intervention to ameliorate the disappointment that actual events bring to bear. That intervention might take the relatively mild form of education (like that considered and rejected in the previous subsection); it might consist of discouraging people from buying some forms of insurance, on the theory that they will adapt to losses; or it might take the more aggressive form of a progressive consumption tax, under which luxury goods would have a higher marginal tax than nonluxury goods (Frank 2000).

To show the deep problems that arise from these difficulties of affective forecasting, I consider two cases—damages for tortious injuries and deterrence against crimes.

Suppose that an accident has occurred. Someone has been injured and has brought a civil complaint against the injurer to recover damages. Suppose, finally, that on the issue of liability the jury has found for the plaintiff and now must consider the extent of the damages that the injurer-defendant owes the victim-plaintiff. If the plaintiff's attorney has shown the jury a "day in the life" video of the extensive health care help that the plaintiff must now have in dressing, eating, going to the bathroom, bathing, and so on, then the jury is likely to award a far higher sum to the plaintiff than the extent of the injuries may warrant. The reason is that individual jurors may be engaging in affective forecasting, asking themselves how they would feel if they had to suffer these injuries and indignities.

In thinking about how they would feel if the injuries happened to them, individual jurors are unlikely to recognize that they (and the plaintiff) would adapt to even these terrible injuries. As a result of these adaptation, their (and her) recovery in well-being is likely to occur within a year of the injury, restoring her state of subjective well-being to its preinjury level (or nearly so). By ignoring the fact of adaptation, jurors may tend to overcompensate tort victims (Bagenstos and Schlanger 2007).

There does not seem to be any straightforward corrective to this tendency to overcompensate tort victims (assuming that is, in fact, what happens). Perhaps artful defense attorneys can find ways to work adaptation into their presentation to the jury

or to the judge, but to do so (“These horrific injuries aren’t really as horrible as you might think”) would seem to be churlish and might, therefore, backfire, leading to even larger awards to plaintiffs. Legislated damages caps or a schedule of damages may work in some contexts (they are administratively effective and send relatively clear signals to potential injurers), but without much deeper consideration of the matter, I am not convinced that they can alleviate the problems that affective forecasting creates for tort liability. Perhaps the most effective responses for potential injurers are to take more precaution *ex ante* or to avoid litigation by settling sooner and more generously than they might otherwise be inclined to do.

Consider now another legal issue involving affective forecasting—criminal deterrence. One of the main developments of the economics of crime and punishment over the last forty years has been the elaboration of a comprehensive theoretical model of criminal decision-making and the empirical testing of that theory’s predictions. The literature is, of course, extensive and full of nuance, but it might be summarized as follows: Criminals are rational calculators; they commit crimes if they perceive the expected benefits to exceed the expected costs and refrain otherwise; the evidence suggests that by raising the expected costs of crime, people have been deterred from committing crime; indeed, since the early 1990s there has been a dramatic fall in violent crime and a slightly less but also large decline in nonviolent crime at the same time that the prison population in the United States has soared.<sup>13</sup>

There has, however, been some dissent from this standard view. These dissents are in part premised on human difficulties with affective forecasting and in part on our well-established myopia with respect to the future (Cooter and Ulen 2012, 495–97; Bronsteen, Buccafusco, and Masur 2009). First, we do not appear to be much influenced by the severity of criminal sentencing. If the punishment increases from, say, two years in prison to three years, the additional year has little effect on deterring criminals, especially the young men who commit most violent crimes. David Lee and Justin McCrary (2005) demonstrated this fact in a remarkable study (see also Levitt 1998). The length of the sentence faced by a person who commits a crime increases sharply on the criminal’s eighteenth birthday. Consequently, the deterrence hypothesis predicts a sharp decrease in crime when juvenile delinquents turn eighteen. Lee and McCrary did a careful statistical analysis of Florida arrest data and found no discontinuity in the probability of committing a crime at the age of majority. So, the longer punishments when the criminal turns eighteen apparently are not deterring them from committing crime.<sup>14</sup> This fact has a simple, powerful implication for an efficiency-minded

<sup>13</sup> In 1980 the US federal and state prison population was approximately 500,000. By 2010 it had risen to almost 2.5 million. Although that population has probably fallen slightly in the past several years, it is nonetheless true that in 2009 the United States had the highest incarceration rate (prisoners per 100,000 population) of any country in the world at 743 per 100,000 (Russia was second, at 577 per 100,000). Additionally, even though the United States accounts for about 5% of world population, it has almost 25% of all the world’s prisoners. Cooter and Ulen 2012, ch. 12.

<sup>14</sup> Nonetheless, Levitt (2004) estimates that approximately one-third of the decline in crime between 1991 and 2004 is attributable to the increase in incarceration between 1980 and 2004.

criminal justice policy: Shortening sentences and redirecting expenditures away from prisons and towards police, which would decrease the severity of the punishment and increase its certainty, would probably deter more crimes at no more expense.

In a similar discontinuity-based study, Keith Chen and Jesse Shapiro looked at whether harsher prison conditions—determined by whether federal prison officials assigned a felon a higher security grade—reduce recidivism, all other things held equal. They found that “[i]nmates housed in higher security levels are no less likely to recidivate than those housed in minimum security; if anything, our estimates suggest that harsher prison conditions lead to more post-release crime” (Chen and Shapiro 2007, 1).

These two studies raise questions about the deterrence aspect of imprisonment (its duration, harshness, and certainty), something that RCT-based law and economics has long assumed and has claimed to have found evidence for.

Paul Robinson and John Darley have argued that criminal law does not deter (2004a and 2004b). Let us be very careful about what the authors claim: They believe that the criminal justice system probably does deter crime, but they are very doubtful that criminal laws deter crime. They want to draw a distinction between such actions as the legislative manipulation of sentence length, which they believe does not have a deterrent effect, and such actions as increasing police patrols or the harshness of prison conditions, which they believe might deter crime (Chen and Shapiro notwithstanding).

The authors base their contention on findings in the behavioral sciences. They write that for criminal law to have a deterrent effect on a potential criminal’s conduct choices, the “following three questions must all be answered in the affirmative:”

1. Does the potential offender know, directly or indirectly, and understand the implications for him, of the law that is meant to influence him? That is, does the potential offender know which actions are criminalized by criminal codes, which actions are required, and which conditions will excuse actions which are otherwise criminal?
2. If he does know, will he bring such understanding to bear on his conduct choices at the moment of making his choices?
3. If he does know the rule and is able to be influenced in his choices, is his perception of his choices such that he is likely to choose compliance with the law rather than commission of the criminal offense? That is, do the perceived costs of non-compliance outweigh the perceived benefits of the criminal action so as to bring about a choice to forgo the criminal action? (Robinson and Darley 2004a, 175)

Robinson and Darley argue that there is evidence that none of these premises is true. First, they report on surveys that they and others have conducted in different states about a limited number of legal rules to ascertain how well a random sample of citizens know prevailing criminal laws. One survey found that in a survey of a “target population” (not the general population) of potential offenders, 18% of respondents had no idea what the sanctions for several crimes would be; 35% said that they did not pay attention to what the sanction would be when considering a crime; and only 22% thought they knew exactly what the

punishment would be. So, the authors conclude that “people rarely know the criminal law rules.”

Robinson and Darley also point out that the overall rate of conviction for crimes is extremely low—approximately 1.3% of all crimes result in a conviction, and the chances of a convicted criminal’s receiving a prison sentence is about 1 in 100 for most offenses; “even the most serious offenses, other than homicide, have conviction rates of single digits.” Many in the general population may not know these facts. Rather, they may believe that the chances of being detected, arrested, and convicted are much higher and are, therefore, deterred from committing crime. But career criminals and their friends and relatives are likely to know how low the conviction and punishment rates really are.

One of the most intriguing points that Robinson and Darley make is that the duration of prison sentences may not have a deterrent effect. They note that people adapt fairly quickly to changed circumstances; for instance, there is evidence that within six months of incarceration prisoners have returned to their preincarceration level of subjective well-being. And there is compelling evidence that in remembering experiences, we all suffer from “duration neglect”—that is, we do not accurately remember the duration of good or bad experiences (Kahneman 2011, 378–81). So, thoughts of imprisonment may deter those of us who have not been “inside” or known anybody who has been imprisoned, but perhaps those who have been imprisoned recall the experience as not as bad as they had anticipated.

## 4 CRITIQUES OF BEHAVIORAL LAW AND ECONOMICS

---

A few economists have urged caution in taking the conclusions of behavioral economics too far before additional work is done (Pesendorfer 2006; Levine 2012), but most economists have accepted that the field has contributed significantly to the understanding and prediction of economic behavior and probably will continue to do so.<sup>15</sup> Within the legal academy, behavioral law and economics has had a very large impact on law and economics. In part this is due to the long-held general suspicion of theorizing within the legal academy and especially to the more recent unhappiness among traditional legal scholars with RCT, which came into legal discourse in the baggage of law and economics.

However, there have been critics of behavioral economics, both within the economics and legal fields. In this section I will, first, discuss how we might think of the anomalies in RCT’s predictions that have been highlighted by behavioral economics as

<sup>15</sup> Parts of the material in this section appear in Ulen, forthcoming. It is, I think, noteworthy that there is, to my knowledge, no textbook in behavioral economics. The field has not, so far, generated a comprehensive and coherent alternative to the standard microeconomic model. Frank (2010) is one of the few microeconomics texts that includes extensive coverage of behavioral results.

a part of the predictable life of a normal science (Cooter 2011). Second, I shall trace the outlines of the various arguments that economists have made in criticizing the behavioral findings of consumer, producer, and investor nonrationality and its potential to disrupt markets. The gist of those criticisms is that even if there are nonrational actors in the marketplace, either rational traders can make money by arbitraging away their mistakes—thereby correcting the market impacts of the irrational traders—or the marketplace has corrective forces that signal nonrational actors about their mistakes in palpable ways, leading to corrections.

Next I will survey the criticisms of behavioral law and economics that have been made by legal scholars, principally those made by Judge Richard Posner and Professors Greg Mitchell and Jon Klick.

I conclude with a brief account of an important philosophical criticism of regulatory intervention based on behavioral insights.

#### **4.1 RCT and Behavioral Economics as Part of the Predictable Course of a Normal Science: “It Takes a Theory to Beat a Theory”**

One criticism that is frequently heard is this: Because behavioral economics has not yet provided a coherent and complete account of human choice, we should stick with RCT, which has provided such a complete account. Or “It takes a theory to beat a theory.” That criticism seems to me to be wrong. Not only is it unhelpful and potentially very damaging; it also ignores the process by which normal science advances.

Any “normal science” has a prevailing “paradigm,” as Thomas Kuhn memorably called it (Kuhn 1996). That paradigm or standard model contains the science’s best current explanation for the real phenomena that the science seeks to understand. Those adept in a given field are familiar with this paradigm or standard model, having learned it as part of their preparation for the profession; their scholarly work largely consists of refining that model, both theoretically and empirically; and an important part of their work is to teach that model to their students, sometimes including a discussion of the shortcomings of that model. Some of those students will train to become adept in the field themselves, demonstrating their abilities in dissertations and subsequent scholarship that extends the standard model or explores some of that model’s shortcomings (although this latter is a riskier strategy for getting a position in the field).

Because the paradigm is, like all theories, a simplification of the perceived complexities of the real-world phenomena that a science seeks to explain and predict, it is inherently incomplete. The theory simply cannot have a one-to-one correspondence to the real phenomena that it seeks to explain. Inconsistencies between theory and reality necessarily exist, both because of the possibility of errors in the theory and because of mismeasurement of data. Much of the valuable scholarship in a field consists of clearing up the gaps in the standard model, explaining seeming inconsistencies, extending

the model to new phenomena (or historical phenomena), and subjecting the model to new data.

Kuhn explains that any normal science has these “anomalies” (what I called inconsistencies between observation and the paradigm). These anomalies can sometimes be incorporated into the standard model. But in other instances these anomalies pile up, making it more and more difficult to amend the standard model so as to accommodate them without the model’s becoming exceedingly complicated and clumsy. At that point, someone typically develops a new paradigm that incorporates all the anomalies that sank the old paradigm but does so in a far less complicated and more elegant manner.<sup>16</sup>

My strong sense is that behavioral economics and RCT are engaged in a struggle like that of the geocentric and heliocentric theories of our planetary system. There have been attempts by RCT to accommodate the anomalies found by Kahneman, Tversky, and others. But my impression is that those accommodations are ad hoc rather than general, as I shall indicate in the following section. I further sense that the weight of the anomalies within RCT is increasing and that a new paradigm is coming.

Having said all that, I think that it is also possible that we shall discover in the future that there are some people for whom and some circumstances for which rational choice is a fine descriptive and predictive theory. In addition, it is possible that we shall discover better ways in which to influence choice architecture (as Thaler and Sunstein helpfully call it) and alternative methods of decision-making (computers serving as diagnostic physicians; the “wisdom of crowds”; and the like) so as to more frequently enjoy the welfare-enhancing benefits of rational choice.

## 4.2 Criticisms of Behavioral Economics by Economists

The rise of behavioral economics has been greeted with mixed emotions by professional economists. On the one hand, the elaboration of the theory of rational choice has been one of the great accomplishment of modern microeconomics and is at the core of a modern graduate education in microeconomics. On the other hand, behavioral economics is a remarkably vibrant and interesting scholarly innovation that has been recognized by the award of a Nobel Prize in Economic Sciences to a psychologist, Daniel Kahneman, in 2002.<sup>17</sup>

<sup>16</sup> Kuhn contends that there are no metrics, other than simplicity and elegance, by which to evaluate succeeding paradigms in a given field. The complicated theory of a geocentric planetary system and the elegant theory of a heliocentric planetary system can both be consistent with all observations. This strikes me as wrong: There are better and worse paradigms—at a minimum, the better paradigm predicts real phenomena better.

<sup>17</sup> Of course, Herbert Simon won the Nobel Prize in 1978 for, among other things, his work on bounded rationality, particularly within business organizations. See Simon’s Nobel Prize lecture (Simon 1978).

In a sense, these two competing accounts of human decision-making are incompatible. That fact has put the profession into a quandary in which many feel that they must choose one or the other of these accounts of human decision-making. And once having done so—the behavioralist is quick to note—an economist, like everyone else, is subject to the confirmation bias (Hastorf and Cantril 1954; MacCoun 1998; O'Brien and Ellsworth 2006) and, therefore, has an incentive to discount information that is favorable to the other view and to exaggerate the importance of information that favors her view.

More significantly, the rise of behavioral economics has induced rational choice theorists to push back. That has involved one of two general lines of criticism of behavioral economics. In the first, the rational choice theorist tries to show that RCT is, in fact, aware of and consistent with the behavioral point. For example, she might note that most people are well aware of their own cognitive shortcomings and seek to avoid being put in situations in which those shortcomings can harm them. Some people know that they have difficulty making a commitment to awaken at a particular time in the morning and, so, place the alarm clock across the room so that they will have to get out of bed to stop the noise. Presumably, they know that once up to walk across the room to shut off the alarm, they had better stay up and get ready to go to work or to their appointment.

The second line of criticism suggests that the behavioral analysis has missed the important fact that decision-makers may learn the value of more rational behavior through experience and, importantly, that the competitive market may be a significant device for inducing more rational behavior. For example, in an important early article on these matters, Gary Becker (1962) argued that even if there are irrational traders in a competitive market for goods or services, rational traders will determine the market price by trading at the margin and by arbitraging away any deviations from marginal cost pricing. As a result, even if there are a large number of irrational traders, Becker demonstrated, demand and supply curves will still have their respective downward- and upward-sloping characteristics.<sup>18</sup>

More recently, John List (2003, 2006) and List and Millimet (2008) have elaborated and extended Becker's argument by using results of field experiments to show that behavioral deviations from the predictions of RCT may be corrected in normal markets.

In the first of his articles, List studied the behavior of professional and ordinary traders at sports card memorabilia shows and at collector pin trading meets. He describes his results as follows:

First, [] I observe a significant endowment effect in the pooled data. Second, I find sharp evidence that suggests market experience matters: across all consumer types,

<sup>18</sup> List and Millimet (2008) confirm Becker's theoretical finding with very carefully designed experiments involving market trades by children between the ages of six and eighteen. "Empirical results generated from observing more than 800 experimental subjects indicate that (i) only about 31% of agents exhibit behavior consistent with rational choice theory, and (ii) market experience facilitates the development of such behavior" (2).



marketlike experience and the magnitude of the endowment effect are inversely related. In addition, with the group of subjects who have intense trading experience (dealers and experienced nondealers), I find that the endowment effect becomes negligible. . . . [T]hese empirical results certainly suggest that individual behavior converges to the neoclassical prediction as market experience intensifies. (List 2003, 42–43)

List recognizes that there are at least two possible explanations for the lessening of the endowment effect among those with significant market experience: (1) market experience itself lessens the effect, or (2) they have, prior to coming to the market, preferences that exclude the endowment effect and, as a result, they trade more often. By returning to the same sports card markets one year after his initial investigations and examining the trading rates for the same subjects who had participated in the prior investigations, he concludes that “market experience significantly attenuates the endowment effect” (List 2003, 43).

This criticism of the endowment effect has been answered persuasively. In particular, there is no endowment effect when goods are held for exchange—such as commercial stock and the sports cards in the dealers’ hands in List’s studies—rather than for use. Otherwise, in buyer-seller relationships the parties’ respective endowment effects would have canceled each other out (see chapter 12 by Korobkin in this volume). Moreover, in a subsequent study, Haigh and List (2005) compared decisions made by undergraduate students and professional futures and options pit traders recruited from the Chicago Board of Trade. Their data suggested that “professional traders exhibit myopic loss aversion to a *greater* extent than students” (523).

In a further series of experiments, List (2006) explored the extent to which prosocial preferences were influential in actual market behavior. (Laboratory and field experiments suggest that prosocial behavior is far more common than RCT would seem to suggest. See chapter 8 by Stout in this volume.) He finds that prosocial behavior is much less common in actual market transactions unless sellers have reputational concerns, which are likely to arise because of repeat dealings with local buyers. He also finds that when quality-monitoring practices are readily and inexpensively available, reputational concerns are more common—that is, that quality-monitoring technology and reputation are complements.

Levitt and List (2007 and 2008) have also had a slightly different line of attack on behavioral economics. Behavioral insights have largely (but not exclusively) come from carefully designed laboratory experiments. Levitt and List have cautioned about drawing inferences about real-world behavior from lab experiments. This is because

behavior in the lab is influenced not just by monetary calculations, but also by at least five other factors: (1) the presence of moral and ethical considerations; (2) the nature *and* extent of scrutiny of one’s actions by others; (3) the context in which the decision is embedded; (4) self-selection of the individuals making the decisions; and (5) the stakes of the game. (Levitt and List 2007, 154)

As an example, they contend that the finding that prosocial behavior is far more common than RCT might be said to claim is, perhaps, a fiction of the lab setting:

[M]any real-world markets operate in ways that make pro-social behavior much less likely. In financial markets, for instance, the stakes are large, actors are highly anonymous, and little concern seems to exist about future analysis of one's behavior. Individuals with strong social preferences are likely to self-select away from these markets, instead hiring agents who lack such preferences to handle their financial dealings.

There is much more bark than bite to these criticisms from economists. Some of them have been answered by experimental and field studies that show the effects of cognitive and judgmental biases in the real world (Camerer 2000; DellaVigna 2009).<sup>19</sup> Many of them are not at all incompatible with the central findings of behavioralists. For instance, it would not surprise anyone familiar with behavioral law to discover that careful experiments find that greater market experience ameliorates some RCT deviations in some people. Nor is there anything controversial about cautioning scholars not to draw inferences too quickly from laboratory experiments about real-world behavior (chapter 5 by Engel in this volume). Anyone familiar with empirical work would second that caution. The answer to some of these criticisms, in short, is to do more and even better scholarship.

### 4.3 Criticisms of Behavioral Law and Economics from the Legal Academy

When comprehensive reexaminations of law and economics through behavioral findings began to appear in the scholarly legal literature (for example in Jolls, Sunstein, and Thaler 1998 and Korobkin and Ulen 2000), they received an enthusiastic reception. In part this enthusiasm arose from the fact that a large number of legal scholars disliked law and economics and saw behavioral studies as a means of attacking law and economics (Kelman 1979). Others welcomed the findings, believing them to be a potentially significant advance in our understanding of how decision-makers might respond to legal rules and standards.

Among those legal scholars who were less enthusiastic about behavioral law and economics was Richard Posner. In a reply to Jolls, Sunstein, and Thaler, Judge Posner (1998) contended that the authors (and the field, generally) used the word "irrational" to describe something that was better described as a mere preference. For example, Jolls, Sunstein, and Thaler described fear of flying as "irrational" because a clear-eyed view of the risks of that mode of travel, by comparison to the risks of alternative modes of travel, suggests that a rational person would not be fearful of flying. Posner suggested calling a "fear of flying" a taste or preference without the pejorative "irrational" attached. Similarly, economics has long taken voting in a democratic election to be "irrational," given the vanishingly small probability that one's vote could influence

<sup>19</sup> I am extremely grateful to Eyal Zamir for making this point and providing these references.

the outcome—that is, that the costs of voting almost certainly exceed the benefits. But Posner suggested that whether to vote may simply be a matter of taste and that a far more interesting research agenda is not to explain why people vote but rather to explore any systematic differences in the desire to vote—for example, to see if the old vote more than the young, the retired more than the unemployed, and so on.

Posner went on to suggest that much of the other examples of irrational behavior given by Jolls, Sunstein, and Thaler are easily incorporated within a rational choice framework. For example, RCT can and has treated bounded rationality or “limited computational power” as simply another constraint (like those of income, wealth, and time) within which rational individuals maximize. Similarly, he suggests that RCT can accommodate hyperbolic discounting as a response to uncertainty and information costs.<sup>20</sup>

Posner further contended that RCT is flexible enough to accommodate other puzzles that behavioral economics has uncovered. For example, rational choice may explain behaving in what appears to be an altruistic manner—a strictly non-self-interested way—as the result of interdependent utilities (that is, that my well-being depends, perhaps in very large part, on my children’s and grandchildren’s well-being).

The one behavioral insight that Posner found genuinely puzzling for RCT is weakness of will (*akrasia*), such as “refusing to keep chocolate in the house because of a fear of not being able to overcome temptation.” Posner contends that the only way in which RCT can accommodate this sort of puzzle is to abandon the view that the self is a unity in favor of the conception of multiple selves. He went further: Each of these multiple selves may be rational (in the sense of having transitive, well-ordered preferences) but that some of the selves’ preferences are inconsistent with those of other selves. The “present self” may prefer, for example, to snack on chocolates today while the “future self” realizes that snacking on chocolates today may lead to health issues in the future and would, therefore, prefer not to indulge in eating today that may be detrimental to well-being in the future. My present self may prefer to luxuriate on the couch today; my future self may prefer that I exercise for thirty minutes today.

There are, naturally, deep issues involved in identifying how many of these multiple selves there are, how they reach an accommodation among themselves to take action today or plan to do something next August. But Posner is surely right in suggesting that

<sup>20</sup> RCT seems to imply that in discounting future costs and benefits to present value, individuals use an idiosyncratic (depending on their tastes for risk and immediate versus delayed gratification) but constant discount rate. So, for example, I generally prefer \$10 today to \$10 in thirty days. The rate at which I would calculate the present discounted value of \$10 in thirty days is the same rate that I would use to compare \$10 in forty-five days to \$10 in fifteen days. However, in many experiments, individuals seem to use different discount rates depending on how distant the reward is and the relative size of the immediate versus the future reward. “Hyperbolic discounting” refers to the empirical observation that people prefer smaller amounts received sooner to larger amounts received later. If, however, people are put to a choice between two distant (in time) payoffs, they apparently prefer the larger. Lee Fennell gives this example: Prudence says that she would prefer \$105 in 366 days to \$100 in 365 days but that she would prefer \$100 today to \$105 tomorrow.

there are these internal conflicts within humans and that a sensible research project is to begin to get at a clearer understanding and set of predictions about human behavior in terms of these multiple selves.

Finally, Posner was bothered by the fact that Jolls, Sunstein, and Thaler and the field of behavioral law and economics did not seem to have any clear predictions about behavior. “It is profoundly unclear what ‘behavioral man’ would do in any given situation.” Indeed, he suggested that such predictive theory as they have “seems perilously close to the abyss of non-falsifiability.” So, when people act rationally, Jolls, Sunstein, and Thaler do not treat this as contradicting bounded willpower. Nor do they have an explanation or prediction about why and when some people resist temptation.

There is much more of interest in Posner’s critique. Of the many fascinating things that he has to say I am particularly intrigued by his suggestion that we should look to evolutionary biology to explain some of the seemingly odd, unconscious behaviors that behavioral economics has identified. For example, the facts that altruism is rewarded and failure to cooperate punished may be due to hardwired accommodations incorporated into human behavior when we lived in small groups.

I do not disagree with the thrust of much of Posner’s critique of Jolls, Sunstein, and Thaler. But I think that the distance between his views and those of someone sympathetic to behavioral law and economics is not nearly so great as might seem to be the case. Thus, I do not think that Posner’s suggestion that investigators look to evolutionary biology for an explanation of hardwired decision-making algorithms that lead us astray in the modern world is inconsistent with anything in behavioral law and economics. Moreover, that suggestion picks up on an important point made by Owen Jones (2001) regarding the difference between being able to affect behavioral changes that are hardwired (the product of relatively slow-moving evolutionary change) and those that are softwired (the product, say, of environment and culture).<sup>21</sup>

Moreover, I think that Posner’s suggestion that behavioral economics is in danger of becoming a “just so” story—much as RCT has become—is a very powerful criticism. One of the early criticisms of RCT was that it was tautological: *Every* behavior, no matter how counterintuitive or loopy-seeming, was the result of some rational choice; it was the investigator’s task to figure out how. So, if people behave rationally and you see a good friend walking through the mall clucking like a chicken and flapping his arms as if they were wings, your presumption should be that there is some rational explanation for this—for example, that he has lost a bet. But a tautology loses explanatory and predictive power because it is always, by definition, true.

Similarly, Posner is correct that behavioral law is as yet incomplete in that it cannot explain and distinguish rational behavior and nonrational behavior, nor why some people are rational with respect to some aspects of their lives and nonrational with regard to others, while others are nonrational most of the time and yet others are rational most of the time.

<sup>21</sup> But see Leiter and Weisberg 2010. I am grateful to Eyal Zamir for this reference.

Greg Mitchell (2000, 2003, and 2009, chapter 7 in this volume) and Mitchell and Jon Klick (2006) have made some important criticisms of behavioral law. Mitchell's chapter in the volume restates and refines those criticisms that he has made earlier. Nonetheless, I want briefly, first, to summarize those earlier criticisms and second, respond to them.

Mitchell's earlier articles made four points against the then-emerging behavioral legal literature. First, he argued that the psychological literature did not "support the bleak and simple portrait of pervasive irrationality painted by these scholars" (Mitchell 2002, 1907). Second, neither the behavioral literature nor the RCT literature provides a complete or fully accurate picture of human decision-making. Both accounts are flawed, and it would be a mistake, therefore, to choose one account over the other as one's guide to describing and prescribing behavior (Mitchell 2003, 72). Third, and most importantly in my judgment, Mitchell notes that "individuals differ greatly in their propensities to act rationally and that situations differ greatly in their propensities to elicit rational behavior from individuals" (Mitchell 2003, 73).<sup>22</sup> Fourth, Mitchell (2009) demonstrates that our first-order thoughts—our initial reactions—may well lead to biased outputs but that they are frequently corrected by our second-order thoughts. That is, we sometimes—perhaps often—stop and think before going with our initial intuition.<sup>23</sup> Mitchell suggests, intriguingly, that law is a "cognitive force operating in the form of both conscious and unconscious second thoughts" (Mitchell 2009, 689) that already prevents the errors and misjudgments of System I thinking from directing our actions inappropriately—as, for example, when it prevents us from giving behavioral effect to our prejudicial first impressions (see also Alter et al. 2007).

I think that all four of these points of criticism are important and deserving of thoughtful responses. I sense that the first and second points are not as strong as the final two. I doubt that anyone who takes behavioral law seriously will quibble with Mitchell's contention that we need to do more work to learn more about individual differences in degrees of cognitive and judgmental errors and contextual effects on decision-making effectiveness. Nonetheless, I think that we already know more than enough about human decision-making foibles to make it inappropriate, to say the least, to premise our descriptions and predictions about behavior on RCT.

Mitchell and Jon Klick (2006) make two additional (closely related but distinct) criticisms of behavioral economics as a basis for policy. The first is that to the extent that governmental policymakers devise law to prevent individuals and organizations from

<sup>22</sup> See also the appendix to that article, at 139–67, in which Mitchell summarizes the psychological literature on how decision-making is affected by such factors as sex differences, cognitive dispositions, cultural differences, developmental differences, the decision-maker's role, and repeated play.

<sup>23</sup> Second thoughts are an example of "metacognition," or "thinking about one's thoughts." The article uses the notion of second thoughts to criticize the contention of the Implicit Association Test (IAT), "which can supposedly detect prejudice and stereotype operating at the unconscious, or implicit, level. . . . This research supposedly reveals that the great majority of Americans implicitly associate many historically disadvantaged groups with negative attributes and historically advantaged groups with positive attributes." Mitchell 2009, 691.

making cognitive errors and misjudgments, they are removing incentives to learn. We learn importantly from our own and others' mistakes. If regulators prevent us from making mistakes, then our learning to do better when not so protected and our motivation to look carefully before following a particular decision-making path are attenuated—to our detriment. The second point draws out an implication of the first point to suggest that paternalistic policies may encourage morally hazardous behavior. That is, if we believe that governmental policies protect us from our own carelessness, then why should we take care? Why should we seek to inform ourselves about, for example, the health consequences of tobacco smoking or of failing to exercise or of being obese?<sup>24</sup>

Mitchell and Klick's criticisms are worth taking seriously—but with the following caveat. At the moment their points are hypotheses about the unintended bad things that might flow from regulation and other policy interventions that take their justification from behavioral findings. There is, however, compelling evidence that bad things happen to individuals and organizations as a result of cognitive and judgmental biases. By contrast, Mitchell and Klick point out some previously unremarked possible costs that ought to be set against the benefits of regulating to minimize the costs of behavioral biases. For their conjectures to have an influence, it must be the case that those costs must be real and that they are greater than the benefits of regulating. There is, to my knowledge, no empirical evidence on these matters. Until that evidence is forthcoming, we should, as always, be cautious and prudent in our regulatory policies premised on behavioral findings, but we should not be chilled from proceeding to regulate when the benefits of doing so are significant.

#### 4.4 A Philosophical Criticism of Regulation Based on Behavioral Economics

A slightly different criticism of intervention into individual decision-making on the basis of behavioral insights comes from the philosophical literature on personal autonomy (Spector 2012). That literature began with Plato's and Aristotle's discussions of the role of self-direction in defining the good life, the life well-lived. Aristotle held that our rational natures should take the lead in directing our lives—in particular, that reason should control our passions and desires.

That literature has, obviously, developed considerably in the succeeding 2,500 years. Nonetheless, the issues sounded by Plato and Aristotle have continued to be at the

<sup>24</sup> Mitchell and Klick (2006, 1662–63) urge that if policies are to be made on the basis of behavioral findings, they be made by legislators, not courts, on the ground that legislatures are better than courts at “consider[ing] competing values and marshal[ing] the evidence relevant to optimal institutional design.” I truly wonder at this. Legislatures are notoriously in the grip of well-funded interest groups, which fact can lead legislators to undervalue general public well-being in favor of the interests of those who bring resources to help the legislators' reelection aspirations.

center of that literature—for example, whether reason is the master of the passions or, as Hume contended, their slave.

The relationship of this literature to the subjects considered in this essay is this: If personal autonomy is an independent value for which individuals strive and that governmental policy ought to foster, then that fact may argue for individuals' being allowed to make their own mistakes, to live their lives under their own direction (terms that require exploration) with protection only against fraud, harm to others, and catastrophic consequences for themselves.

We can even approximate this feeling of autonomous well-being today with the help of pharmaceuticals that relieve us of the anxiety and stress of life. Does the notion of an independent, self-directed life mean a life in which the individual experiences all the highs and lows of life unmediated by drugs?

## 5 CONCLUSION

---

When these matters are all taken into account, how does behavioral law fare? My belief is that behavioral law is one of the most important developments—and probably *the* most important—in legal scholarship of the modern era. Even if pre-1980 legal scholarship did not know of RCT, it had, nonetheless, enthusiastically embraced the view that law addressed the “reasonable man or woman.” There is now a great deal of evidence to warrant being skeptical of the proposition that reason is directing our own steps or those of our fellow human beings. There is, clearly, a cost to this fact: People are not as well-off as they would be if they behaved more reasonably. Law—by taking due account of these predictable, routine deviations from rationality—can better influence behavior to realize both social goals and to help individuals better enhance their well-being. Clearly we do not yet know as much as we will or should about human decision-making, and we are novices at figuring out when and how to intervene in individual decision-making on behavioral grounds. It is true that we should be cautious and prudent in intervening. Nonetheless, we do know enough already to know that premising our analyses on the assumption that human beings are always rational decision-makers is a mistake, both analytically and practically.

## ACKNOWLEDGMENTS

---

I am grateful to Eyal Zamir, Doron Teichman, and Avishalom Tor for their invitation to participate in this *Handbook* and a conference at the University of Notre Dame Law School with the other participants. The discussion at that conference was extremely helpful in the revisions. I want particularly to thank Eyal and Doron for their excellent, insightful comments on an earlier draft of this chapter.



My friend and colleague Arden Rowell had planned to join me in writing this essay. She and I spent many hours discussing behavioral law and economics in preparation for the project. Unfortunately, at the last moment and for health reasons Arden was unable to continue her work on this essay. I am extremely grateful to her for her insights and sorry that readers will not have the benefit of those insights in this essay. I have used the word “we” to describe conclusions or ideas that we had discussed and agreed upon. I use “I” to cover those conclusions or ideas that are mine.

## REFERENCES

- Alter, Adam L., Nicholas Epley, Rebecca N. Eyre, and Daniel M. Oppenheimer. 2007. Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General* 136: 569–76.
- Bagenstos, Samuel R., and Margo Schlanger. 2007. Hedonic damages, hedonic adaptation, and disability. *Vanderbilt Law Review* 60: 745–97.
- Barberis, Nicholas C. 2013. Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Literature* 27: 173–96.
- Becker, Gary S. 1962. Irrational behavior and economic theory. *Journal of Political Economy* 70: 1–13.
- Ben-Shahar, Omri, and Carl E. Schneider. 2011. The failure of mandated disclosure. *University of Pennsylvania Law Review* 159: 647–749.
- Blumenthal, Jeremy A. 2005. Law and the emotions: The problems of affective forecasting. *Indiana Law Journal* 80: 155–238.
- Brickman, Philip, Dan Coates, and Ronnie Janoff-Bulman. 1978. Lottery winners and accident victims: Is happiness relative? *Journal of Personality & Social Psychology* 36: 917–27.
- Bronsteen, John, Christopher J. Buccafusco, and Jonathan S. Masur. 2009. Happiness and punishment. *University of Chicago Law Review* 76: 1037–81.
- Bronsteen, John, Christopher J. Buccafusco, and Jonathan S. Masur. 2010. Welfare as happiness. *Georgetown Law Journal* 98: 1583–641.
- Bubb, Ryan, and Richard Pildes. 2014. How behavioral economics trims its sails and why. *Harvard Law Review* 127: 1593–1678.
- Camerer, Colin F. 2000. Prospect theory in the wild: Evidence from the field. In Daniel Kahneman and Amos Tversky, eds., *Choice, Values, and Frames*, 288–300. Cambridge: Cambridge University Press.
- Carlin, Bruce Ian, and David T. Robinson. 2012. Financial education and timely decision support: Lessons from Junior Achievement. *American Economic Review* 102: 305–8.
- Centers for Disease Control and Prevention, United States. 2013. Overweight and obese: Adult obesity facts: Obesity is common, serious, and costly. <http://www.cdc.gov/obesity/data/adult.html>.
- Chen, M. Keith, and Jesse M. Shapiro. 2007. Do harsher prison conditions reduce recidivism? A discontinuity-based approach. *American Law and Economics Review* 9: 1–29.
- Clark, Robert L., Melinda Sandler Morrill, and Steven G. Allen. 2012. Effectiveness of employer-provided financial information: Hiring to retiring. *American Economic Review* 102: 314–18.

- Conly, Sarah. 2013. *Against Autonomy: Justifying Coercive Paternalism*. Cambridge: Cambridge University Press.
- Cooter, Robert D. 2011. Maturing into normal science: The effect of empirical legal studies on law and economics. *University of Illinois Law Review* 2011: 1475–83.
- Cooter, Robert D., and Thomas S. Ulen. 2012. *Law and Economics*. 6th ed. Upper Saddle River, NJ: Prentice Hall.
- DellaVigna, Stefano. 2009. Psychology and economics: Evidence from the field. *Journal of Economic Literature* 47: 315–72.
- Diener, Ed, & Martin E. P. Seligman. 2004. Beyond money: Toward an economy of well-being. *Psychological Science in the Public Interest* 5: 1–31.
- Dunn, Elizabeth W., Daniel T. Gilbert, and Timothy D. Watson. 2011. If money doesn't make you happy, then you're probably not spending it right. *Journal of Consumer Psychology* 21: 115–125.
- Dunn, Elizabeth W., and Michael Norton. 2013. *Happy Money: The Science of Smarter Spending*. New York: Simon & Schuster.
- The Economist*. 2013. Video games and financial literacy: Blood bars and debt bunnies. February 2, 63.
- Fernandes, Daniel, John G. Lynch Jr., and Richard Netemeyer. 2013. Financial literacy, financial education, and downstream financial behaviors. *Management Science*. <http://pubsonline.informs.org/doi/abs/10.1287/mnsc.2013.1849?journalCode=mnsc>.
- Frank, Robert H. 2000. *Luxury Fever: Money and Happiness in an Era of Excess*. Princeton, NJ: Princeton University Press.
- Frank, Robert H. 2010. *Microeconomics and Behavior*. 8th ed. Burr Ridge, IL: McGraw-Hill Higher Education.
- Gilbert, Daniel T. 2006. *Stumbling on Happiness*. New York: Knopf.
- Haigh, Michael S., and John A. List. 2005. Do professional traders exhibit myopic loss aversion? An experimental analysis. *Journal of Finance* 60: 523–34.
- Hastorf, Albert, and Hadley Cantril. 1954. They saw a game: A case study. *Journal of Abnormal & Social Psychology* 49: 129–34.
- Jimenez, Dalié, D. James Greiner, Lois R. Lupica, and Rebecca L. Sandefur. 2013. Improving the lives of individuals in financial distress using a randomized control trial: A research and clinical approach. *Georgetown Journal on Poverty Law and Policy* 20: 449–77.
- Johnson, Eric J., and Daniel Goldstein. 2003. Do defaults save lives? *Science* 302: 1338–39.
- Johnson, M. R., K. H. Meyer, J. Haft, D. Kinder, S. A. Webber, and D. B. Dyke. 2010. Heart transplantation in the United States, 1999–2008. *American Journal of Transplantation* 10: 1035–46.
- Jolls, Christine, Cass R. Sunstein, and Richard H. Thaler. 1998. A behavioral approach to law and economics. *Stanford Law Review* 50: 1471–550.
- Jones, Owen. 2001. Time-shifted rationality and the law of law's leverage: Behavioral economics meets behavioral biology. *Northwestern University Law Review* 95: 1141–205.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263–92.
- Kahneman, Daniel, Alan B. Krueger, David Schkade, Norbert Schwarz, and Arthur A. Stone. 2006. Would you be happier if you were richer? A focusing illusion. *Science* 312: 1908–10.
- Kelman, Mark. 1979. Consumption theory, production theory, and ideology in the Coase Theorem. *Southern California Law Review* 52: 669–98.

- Korobkin, Russell B., and Thomas S. Ulen. 2000. Law and behavioral science: Removing the rationality assumption from law and economics. *California Law Review* 88: 1051–144.
- Kuhn, Thomas S. 1996. *The Structure of Scientific Revolutions*. 3rd ed. Chicago: University of Chicago Press.
- Lee, David, and Justin McCrary. 2005. Crime, punishment, and myopia. NBER Working Paper No. W11491.
- Leiter, Brian, and Michael Weisberg. 2010. Why evolutionary biology is (so far) irrelevant to legal regulation. *Law and Philosophy* 20: 31–74.
- Levine, David K. 2012. *Is Behavioral Economics Doomed?: The Ordinary versus the Extraordinary*. Cambridge: Open Book Publishers.
- Levitin, Adam. 2013. The Consumer Financial Protection Bureau: An introduction. <http://www.ssrn.com>.
- Levitt, Steven D. 1998. Juvenile crime and punishment. *Journal of Political Economy* 106: 1156–85.
- Levitt, Steven D. 2004. Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not. *Journal of Economic Perspectives* 18: 163–90.
- Levitt, Steven D., and John A. List. 2007. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21: 153–74.
- Levitt, Steven D., and John A. List. 2008. *Homo economicus* evolves. *Science* 319: 909–10.
- List, John A. 2003. Does market experience eliminate market anomalies? *Quarterly Journal of Economics* 118: 41–71.
- List, John A. 2006. The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *Journal of Political Economy* 114: 1–37.
- List, John A., and Daniel L. Millimet. 2008. The market: Catalyst for rationality and filter of irrationality. *Berkeley Electronic Journal of Economic Analysis and Policy* 8: 1–53.
- Lusardi, Annamaria, Pierre-Carl Michaud, and Olivia S. Mitchell. 2012. Optimal financial knowledge and wealth inequality. <http://www.ssrn.com>.
- Lusardi, Annamaria, and Olivia S. Mitchell. 2009. Financial literacy: Evidence and implications for financial education. TIAA-CREF Institute, *Trends and Issues*, May, 1–10.
- MacCoun, Robert J. 1998. Bias in the interpretation and use of research results. *Annual Review of Psychology* 49: 259–87.
- Mitchell, Gregory. 2002. Taking behavioralism too seriously? The unwarranted pessimism of the new behavioral analysis of law. *William and Mary Law Review* 43: 1907–2021.
- Mitchell, Gregory. 2003. Why law and economics' perfect rationality should not be traded for behavioral law and economics' equal incompetence. *Georgetown Law Journal* 91: 67–167.
- Mitchell, Gregory. 2009. Second thoughts. *McGeorge Law Review* 40: 687–722.
- Mitchell, Gregory, and Jonathan Klick. 2006. Government regulation of irrationality: Moral and cognitive hazards. *Minnesota Law Review* 90: 1621–63.
- O'Brien, Barbara, and Phoebe C. Ellsworth. 2006. Confirmation bias in criminal investigations. <http://www.ssrn.com>.
- Orentlicher, David. 2009. Presumed consent to organ donation: Its rise and fall in the United States. *Rutgers Law Review* 61: 295–331.
- Pesendorfer, Wolfgang. 2006. Behavioral economics comes of age: A review essay of *Advances in Behavioral Economics*. *Journal of Economic Literature* 44: 712–21.
- Posner, Richard A. 1998. Rational choice, behavioral economics, and the law. *Stanford Law Review* 50: 1551–75.

- Riker, William H. 1986. *The Art of Political Manipulation*. New Haven, CT: Yale University Press.
- Robinson, Paul H., and John M. Darley. 2004a. Does criminal law deter? A behavioral science investigation. *Oxford Journal Legal Studies* 24: 173–205.
- Robinson, Paul H., and John M. Darley. 2004b. The role of deterrence in the formulation of criminal law rules: At its worst when doing its best. *Georgetown Law Journal* 91: 941–1002.
- Schkade, David A., and Daniel Kahneman. 1998. Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science* 9: 340–46.
- Schwartz Center for Economic Policy Analysis, Department of Economics, The New School. 2012. Retirement account balances by income: Even the highest earners don't have enough. <http://www.economicpolicyresearch.org/guaranteeing-retirement-income/528-retirement-account-balances-by-income-even-the-highest-earners-dont-have-enough.html>.
- Seligman, Martin E. P. 2011. *Flourish: A Visionary New Understanding of Happiness and Well-Being*. New York: Free Press.
- Shafir, Eldar, ed. 2013. *The Behavioral Foundations of Public Policy*. Princeton, NJ: Princeton University Press.
- Simon, Herbert A. 1978. Rational decision making in business organizations. *American Economic Review* 69: 493–513.
- Social Security Administration, United States. 2013. Social security basic facts. <http://www.ssa.gov/pressoffice/basicfact.htm>.
- Spector, Horacio. 2012. Autonomy. In Gerald F. Gauss and Fred D'Agostino, eds., *The Routledge Companion to Social and Political Philosophy*. New York: Routledge: 573–84.
- Sunstein, Cass R., and Richard H. Thaler. 2003. Libertarian paternalism is not an oxymoron. *University of Chicago Law Review* 70: 1159–202.
- Thaler, Richard H., and Shlomo Benartzi. 2004. Save More Tomorrow: Using behavioral economics to increase employee savings. *Journal of Political Economy* 112.1, Part 2: S164–S187.
- Tversky, Amos, and Daniel Kahneman. 1981. The framing of choices and the psychology of choice. *Science* 211: 453–58.
- Ulen, Thomas S. Forthcoming. Behavioral law and economics: Law, policy, and science. *Supreme Court Economic Review*.
- Willis, Lauren E. 2011. The financial education fallacy. *American Economic Review* 101: 429–34.
- Willis, Lauren E. 2014. When nudges fail: Slippery defaults. *University of Chicago Law Review* 80: 1155–1229.
- Zamir, Eyal. 1998. The efficiency of paternalism. *Virginia Law Review* 84: 229–86.
- Zamir, Eyal. 2012. Loss aversion and the law. *Vanderbilt Law Review* 65: 829–94.
- Zamir, Eyal, and Barak Medina. 2010. *Law, Economics, and Morality*. Oxford: Oxford University Press.

## CHAPTER 5

---

# BEHAVIORAL LAW AND ECONOMICS

## *Empirical Methods*

---

CHRISTOPH ENGEL

### 1 THE LANDSCAPE

---

EVENTUALLY, all law is behavioral (for a predecessor paper see Tor 2008). In a forward-looking perspective, this is obvious. In this perspective, law is a governance tool (see Posner 2011). All the law is potentially able to achieve is a change in human behavior. Of course, not all lawyers agree with this mission statement. In a backward-looking perspective, the critical issue is attribution. The minimum requirement for attribution is possible cause. A person has possibly acted in some way, and this has possibly led to detriment. Whether the person has actually acted is a behavioral question. Usually the law is not content with mere causation and requires intent or negligence. That makes the behavioral dimension even more prominent. Of course the person may be a legal person, for example, a firm. Then the additional complexity resulting from the presence of a corporate actor must be handled (Engel 2010). Still the core of the matter is how the law should react to some course of behavior. Finally, in a deontological perspective, the law emanates from and impacts on moral intuitions prevalent in society. Whether individuals truly hold a purported moral intuition, and how these intuitions and explicit legal rules interact, is another behavioral issue (Zamir and Medina 2011).

In two areas of law, a behavioral perspective has a long tradition. Law and psychology scholars for long have studied psychology in the courtroom. Criminologists have long tried to understand why there is crime and how it can be mitigated by criminal law. Behavioral law and economics is a more recent phenomenon (see the programmatic book by Sunstein 2000). As law and economics in general, it adopts an individualistic perspective. Legal rules are understood (and often explicitly modeled) as changes in the opportunity structure. The overall research question is explaining actions of individuals by their reaction to this opportunity structure. Behavioral

law and economics deviates from traditional law and economics by the assumptions about the driving forces of individual action. In the standard framework, individuals have well-defined and well-behaved preferences, they know everybody else's preferences, and they dispose of unlimited cognitive abilities. In line with the behavioral approach to all economics, behavioral law and economics relaxes these assumptions. It allows for richer utility functions, and in particular for social preferences. Actors are no longer assumed to exclusively care about their own well-being. Behavioral law and economics is also open to a panoply of cognitive effects, and investigates in which ways legal institutions reflect or mitigate them. Behavioral law and economics also adopts a broader topical scope than the two older behavioral traditions. It focuses on the behavioral foundations of all law, and in particular of private and public law.

Most behavioral lawyers are consumers of scientific evidence. They, for instance, capitalize on the heuristics and biases literature in cognitive psychology (for a summary account, Kahneman and Tversky 2000) and investigate its implications for the interpretation and the design of the law. Most of these contributions come out in the law reviews (e.g., Rachlinski 2011). Sometimes a broader theme leads to a monograph. I have for instance argued that, taking the existing behavioral evidence into account, human behavior seems utterly unpredictable. This points to a neglected purpose of legal intervention. Rather than taming socially undesirable motives, through institutional intervention the law makes behavior predictable and thereby social interaction meaningful (Engel 2005).

In many dimensions, the existing behavioral evidence is rich and differentiated. Lawyers have no reason to reinvent the wheel. Yet not so rarely a lawyer is unable to spot the evidence she needs for her normative business. There are two main reasons why a neighboring discipline has not delivered. The lawyer may want to know whether a general effect also holds under the specific conditions the law aims to address. Or the specific behavioral effect the lawyer suspects to be critical has just not been on the screen of psychologists or economists. In such situations, legal scholars would ideally wish to generate fresh behavioral evidence. This of course requires expertise in empirical methods, or collaboration with colleagues who have this training. It is a happy coincidence that the empirical legal movement is growing so rapidly these days. Many believe that the movement has chiefly been ignited by the availability of big data. Not all empirical law is behavioral. But a substantial fraction is, and it benefits from mounting interest in the empirical foundations of doctrinal or legal policy argument.

Sometimes, trained lawyers publish in the psychology (e.g., Simon et al. 2004) or in the economics journals (e.g., Zeiler and Plott 2005). Sometimes, fresh behavioral evidence is the core of a law review article (e.g., Buccafusco and Sprigman 2011). Yet the typical outlet for lawyers' efforts at generating new behavioral evidence is the peer-reviewed legal journals. If the paper is written in the psychological paradigm, and uses the methods prevalent in psychology, there are two specialized journals, *Psychology*, *Public Policy & Law*, and *Law & Human Behavior*. The former has a slightly broader focus and publishes papers from all subfields of law. It is particularly interested in contributions with direct policy relevance. The latter journal specializes in the two



intersections between law and behavioral research with the longest tradition: criminology and forensic psychology. In principle, all of the many criminology journals are open to behavioral evidence on crime and criminal law. This in particular holds for the flagship journal *Criminology*. Two journals are particularly relevant since they focus on one empirical methodology, the *Journal of Experimental Criminology* and the *Journal of Quantitative Criminology*.

These specialized outlets notwithstanding, scientific empirical evidence is closely tied to law and economics. All the relevant peer-reviewed journals also publish behavioral evidence, the *Journal of Law and Economics*, the *Journal of Law, Economics and Organization*, the *Journal of Legal Studies*, the *American Law and Economics Review*, the *International Review of Law and Economics*, and the *Review of Law and Economics*. The same holds for the newly founded *Journal of Legal Analysis*. Yet in all these journals, empirical contributions compete with theory and policy papers. This is different with the *Journal of Empirical Legal Studies* (hereinafter *JELS*). The journal is entirely devoted to empirical contributions. In the following paragraphs, I focus on this journal since this generates the most accurate portrayal of the emerging discipline.

The journal was founded in 2004. In the first nine years of its existence, it published a total of 227 articles. Seventy-seven can be classified as behavioral.<sup>1</sup> This makes 33.92% of all publications. Figure 5.1 shows that, a slight drop in 2012 notwithstanding, the number of behavioral publications has been steadily growing over time, in both absolute terms and relative to the total number of publications in the respective year.

Thirty-two behavioral papers deal with a question of private law. Twenty-seven cover an issue from criminal law. Public law has only attracted five publications. The remaining 13 papers are not focused on one subdiscipline of law in particular.

Empirical scholars have been much more interested in motivation (57 publications) than in cognition (20 papers).

Publications are strongly US-centric. Sixty-seven papers either address a question from US law, or they use US data or, in an experiment, US subjects. There are three experimental papers from Germany, two papers from the UK, and one from Australia, Israel, and Taiwan each. Finally, two papers compare different jurisdictions.

In the following, I will use the *JELS* data to describe and analyze the methodological variance (section 2). By way of illustration, I will also glance at publications from the remaining peer-reviewed legal journals. In conclusion, I will sketch paths for future methodological development (section 3).

<sup>1</sup> I have coded all papers as behavioral that explore an aspect of motivation or cognition, from whatever conceptual or methodological perspective. The dataset coding these publications is available upon request.



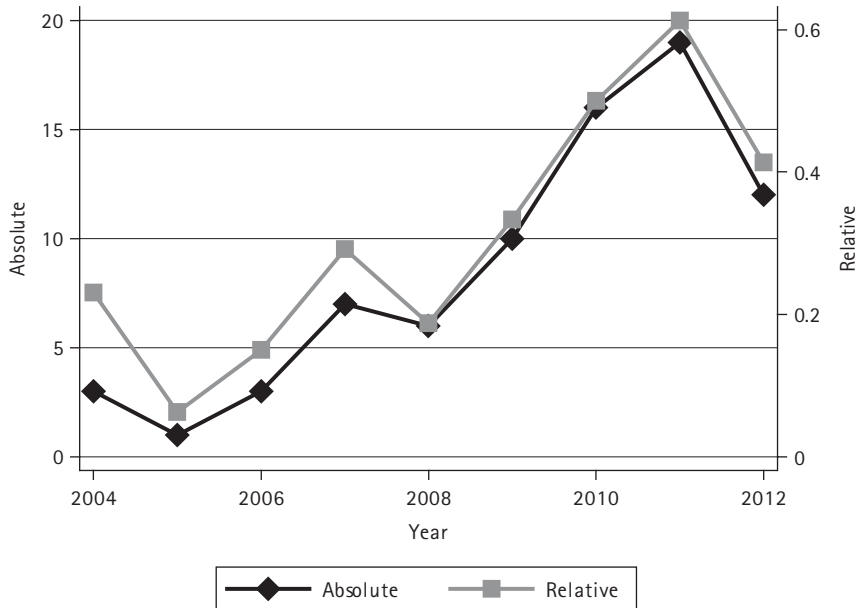


FIGURE 5.1 Behavioral Evidence in the *Journal of Empirical Legal Studies*

## 2 COMPETING METHODOLOGIES

### 2.1 Choice of Method

Method matters. Let me illustrate this with one of the prominent claims of law and economics theory. Traditionally, tort has been constructed as a technology for restoration. The tortfeasor has intruded into the victim's sphere without entitlement, and has caused harm. The victim sues the tortfeasor for redress. The court ruling is meant to make the victim whole. Law and economics scholars object: this construction neglects that rational would-be tortfeasors anticipate the intervention and adjust their behavior. In this forward-looking perspective, tort liability deters socially undesired behavior, much like a criminal sanction or the intervention of a public authority. Quite a few lawyers have been skeptical whether liability is indeed a powerful deterrent. There have been multiple empirical attempts to measure the effect, and the evidence has been mixed (Schwartz 1994).

Recently, three different empirical studies have tried to settle the issue. One study worked with field data. If tort liability is more severe, it should induce would-be tortfeasors to be more careful. This should reduce the overall level of harm. The study used an indirect approach to measure the effect. It tested the following hypothesis: in those US states with stricter rules on medical malpractice, newborns should be in better health. The study did not find a significant association between the two (Yang et al. 2012).

The second study used the methodology prevalent in psychology. First-year law students were exposed to a series of hypotheticals, all of which involved some form of illegal behavior. Between subjects, the authors manipulated the legal regime. One of those regimes was tort liability. Students were asked to indicate how likely they were to engage in one of these activities, given the legal regime in question. Tort liability had little effect. For many scenarios, the proclivity to engage in illegal activity was not significantly different from the condition where no legal rule was mentioned (Cardi et al. 2012).

The third study used the methodology developed in experimental economics, that is, a lab experiment. Participants were exposed to a four-person dilemma. If a participant behaved selfishly, she imposed damage on the remaining group members. The experiment manipulated the certainty and the severity of the right to claim redress. If redress was sufficiently certain, if it was sufficiently severe, and if the threat of compensation had a sufficiently high expected value, the dilemma did not disappear, but it no longer deteriorated over time. Tort liability thus stopped the downward trend of contributions to the joint project that is otherwise typical for these experiments. With this qualification, the experiment found a deterrent effect of tort (Eisenberg and Engel forthcoming, 2014).

Of course, these three studies differed by more than just methodology. It could well be that doctors are less sensitive to the severity of tort liability than students. It could be that professionals are less sensitive than ordinary people. It could be that doctors care less because they are insured. It could be that severity matters for intentional tort but not for negligent tort. It could be that severity matters if the tortious act simultaneously harms multiple victims, but not in a one-to-one relationship. Yet it could also be that one needs a method where individuals actually feel the pecuniary loss of paying damages to see the governance effect of tort.

Behavioral legal scholars have used a considerable plurality of empirical methods. Figure 5.2 shows that two methods are most prominent: field data and vignette studies. Currently, surveys are less visible in *JELS*. Lab experiments have always been least popular.

Any choice of empirical method is a trade-off between external and internal validity. A result is externally valid if there is a good match between the data and the social phenomenon the researcher intends to explain or predict. Seemingly with observational data, external validity is beyond doubt. One directly studies the phenomenon one wants to explain. Yet behavioral researchers have no reason to expect natural laws. They may at best find typical patterns. Ideally they also gain a sense of relative frequency, and of robustness to contextual variation. It therefore would not be meaningful to hunt for the one observation that disproves the claim. Behavioral researchers take it for granted that such exceptions exist. They are content with delineating the conditions under which an effect is normally present.

External validity would still not be an issue if researchers could simply observe the total population. They could then map out the framework conditions of the effect simply by studying it under all possible conditions. Yet behavioral research almost never

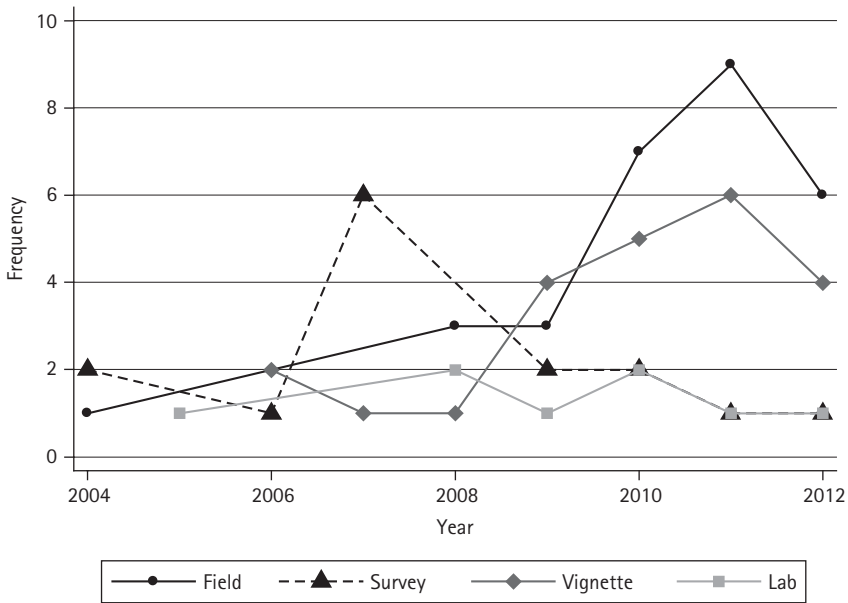


FIGURE 5.2 Empirical Methodologies in the *Journal of Empirical Legal Studies*

observes a total population. Probably the empirical work on decision-making in the US Supreme Court comes closest. But even there, many steps of the decision-making process remain confidential. More importantly researchers want to predict future decisions by the Court. Even if the composition of the Court remains stable, this still requires extrapolation from the past to the future. Normally, behavioral researchers also want to extrapolate from the sample they have observed to the population they want to explain.

For such extrapolation to be legitimate, researchers must be sufficiently confident that the effect they have observed is characteristic for the population. Assessing this confidence is the purpose of statistical analysis. Technically, the researcher compares some characteristic feature of her observation with some well-defined counterclaim. She, for instance, expects that the requirement to justify administrative orders in writing leads to a greater willingness to abide by the order. Let's assume there are two otherwise similar administrative decisions. The law obliges the authority to give written reasons in the first domain, but not in the second. Let us further assume that the researcher finds sufficient qualitative evidence to support the claim that the number of appeals is a reasonable proxy for the willingness to abide by the order just because it is in force. On these assumptions, the researcher could count the number of appeals in both domains. Her statistic would inform her whether this difference is so large that it is very unlikely to result from random variation.

In research practice, this test is entrusted to a statistical package. Researchers usually only report that they have found a significant effect. According to the convention in the social sciences, this statement is warranted if the probability that the difference results from randomness is below 5%. What easily gets lost, though, is that this procedure

requires an explicit *ex ante* hypothesis. The hypothesis is surprisingly often not made explicit. It was missing in 29 of the 77 empirical behavioral papers in *JELS*.

Even if significance is established, this only shows that one phenomenon is associated with another. All one knows is correlation. Typically this is not enough for the normative legal research question. There are three classic topics. Legal scholars want to understand whether there is reason for legal intervention. They want to know whether some legal rule is likely to improve the situation. And they want to assess the quality of some process for rule generation or rule application. All these research questions require the separation of cause and effect. Because consumers are misled by some marketing technique, this technique should be banned. Because individuals understand probability information better if it is presented in the form of natural frequencies, a rule that obliges insurance companies to use this format helps individuals make better choices. Because individual consumers are very unlikely to sue a company for the use of a detrimental standard form contract, class action advances consumer rights.

Empirical methods are on a continuum. Field evidence has the key advantage that one directly studies the phenomenon one wants to understand. Yet with field data, identification is notoriously problematic. The more the researcher takes identification seriously, the more she is at the mercy of unanticipated natural variation. On the other end of the spectrum are lab experiments. They artificially generate an environment, and randomly expose participants to different regimes. If the experiment is well designed, these regimes differ by just one feature. If there is a treatment effect, it must result from this manipulation. Causation is not an issue. But experimenters pay a price. Of necessity, what they study is only analogous to what they want to understand. External validity almost always is an issue. Surveys are one step more artificial than field data. They are handed out to those whose behavior one wants to understand. But the set of questions is designed, not naturally occurring. Vignette studies share many features of lab experiments. They differ by being hypothetical. Participants are asked to assess scenarios, rather than make decisions that matter. These scenarios typically tell a real-life story, and thereby admit considerably more context. Surveys and vignettes are thus in the middle between both extremes. In the following, the power and the limitations of these competing empirical methods are illustrated.

## 2.2 Field Evidence

Behavioral researchers are interested in motivation and cognition. These are difficult to observe directly. It is therefore not obvious how observational data helps answer behavioral research questions. Three very different publications illustrate options.

A first study exploits the fact that juries have to take a joint decision. In preparation, juries deliberate. The state of Arizona considered allowing juries to discuss the evidence during trial. To learn more about the effects of such a reform, the Arizona Supreme Court sanctioned videotaping of randomly selected juries. Researchers exploited the opportunity to test whether jury members were influenced by information that had not

been officially introduced into court procedure, like signals for a witness being wealthy. The study found that this was indeed the case, but that such “offstage” information had little impact on later jury deliberations about guilt (Rose et al. 2010).

Judges should treat all defendants on the merits of their cases. This norm implies that the defendant’s ethnicity should not matter per se. In principle, ethnic bias is difficult to identify since ethnicity is notoriously correlated with the merits of many cases. If one ethnic group is more violent than another, members of this group should be punished more often for violent crime. Yet the individual defendant should not be more likely to be convicted just because of his ethnic background. A study used the fact that, in Israel, over the weekend, suspects are randomly assigned to judges for setting bail. It turned out that Jewish judges were more likely not to detain Jewish than Arab suspects, while it was the other way round for Arab judges (Gazal-Ayal and Sulitzeanu-Kenan 2010).

Many diseases are not exclusively the result of factors beyond the individual’s control, like genetics or accident. Attempts at helping the needy may therefore have the counterproductive effect of reducing self-control. In principle, this offsetting effect is difficult to show since almost all diseases have multiple causes, at least potentially. A study uses legal reform as a source of variation. Some US states have made it mandatory that health insurance plans cover diabetes treatment. The study uses body mass index as a proxy for self-control. Arguably, if diabetics eat more, they increase diabetes problems. The study finds that the gap in the body mass index between diabetics and nondiabetics increased in states that had made coverage for diabetes mandatory (Klick and Stratmann 2007).

Establishing causality is difficult with observational data for two major reasons. Significant correlation may result from reverse causality. Because consumers know that the legal order cares, they stop being vigilant. Significant correlation may also result from an omitted variable that explains both the presumed cause and the presumed effect. In a classic study by Leamer (1983), the true cause for both the choice of a marketing technique and its effect is a lack of literacy in one group of consumers.

Econometricians have developed a whole panoply of techniques to solve such endogeneity problems (for an excellent introduction see Blundell and Costa Dias 2009). Essentially they all rest on some form of quasi-random variation. Empirical legal scholarship on behavioral issues is very differently sensitive to this identification problem. Thirty-two of 77 papers published in *JELS*, and 12 of the 29 studies working with observational data, do not deal with the identification problem at all. Those observational studies that address identification use many different approaches.

The most congenial approach to empirical legal studies is the difference in difference estimator. It is used by six of the 29 relevant publications in the *JELS* (a good example is Frakes 2012). For example, one legislature has changed a critical rule while another legislature has not. This permits identification if the two jurisdictions are sufficiently comparable and if the rule change was the only major difference in development between them. The researcher may then compare the change in the outcome variable before and after the rule change with the change in the same outcome variable in the jurisdiction where rules have not changed. If there is a significant difference between these two

changes in outcomes, it is caused by the change in rules. This procedure has the advantage of cleaning the data of unobserved effects that are idiosyncratic to each jurisdiction. Since the procedure only looks at changes, idiosyncrasies are immaterial as long as they do not change over time, and to the extent that they do not interact with the rule change.

The classic response of econometricians to endogeneity problems is instrumentation. One tries to find an additional variable that is sufficiently correlated with the endogenous explanatory variable, but uncorrelated with the dependent variable. In the most straightforward application, in a first step one explains the potentially endogenous variable by the additional exogenous variable. In the equation explaining the dependent variable one replaces the endogenous variable by the predicted values from the first estimation step. Essentially one now explains the dependent variable by that portion of the independent variable that is cleaned of reverse causation or the omitted variable problem.

A single behavioral paper published in *JELS* uses instrumentation (Brinig and Garnett 2012). The paper investigates one facet of the relationship between religion and crime. The paper hypothesizes that the presence of a Catholic elementary school creates social capital that, in turn, reduces crime. Obviously there could be reverse causality. Catholic schools could be closed because the community deteriorates. Therefore correlation per se is not informative. The authors instrument school closure with irregularities in the parish. Arguably such irregularities are not correlated with social capital in the respective community. They find the effect of religion they were expecting to find.

Occasionally, the law generates randomness for reasons unrelated with research. For instance, in some branches of the US Court of Appeals, judges are randomly assigned to cases. This makes it possible to test for bias resulting from proximity of a judge to the political party of the president by whom she has been appointed. Such bias may indeed be established (Hall 2010).

### 2.3 Survey

Observational data leave the phenomenon one wants to understand completely untouched. Surveys are slightly more intrusive. Participants are interviewed, maybe in writing or online, on their knowledge, understanding, attitude, judgment, or choices. Usually, one and the same participant answers a whole battery of questions. Typically questionnaires are handed out to members of that same population whose behavior one wants to understand. One such study was interested in offenders with mental disorders falsely confessing or falsely pleading guilty (Redlich, Summers, and Hoover 2010). To that end, interviews with detainees in six different institutions were conducted. A single institution significantly differed from the remaining institutions, so that the main findings could be replicated five times. Although participation was not random, this procedure makes it very unlikely that the effects result from selecting atypical cases. Members of minorities, those with a longer criminal career, and

those more severely mentally ill were significantly more likely to report that they had wrongly confessed or pled guilty. The paper discusses the limitations inherent in self-report. It does not address the identification problem: did participants plead guilty because they were mentally ill, or did they become mentally ill because they pleaded guilty?

## 2.4 Vignette

The standard method in law and psychology is a vignette study. Participants are presented with a hypothetical scenario, and they are asked how they would behave themselves were they in that situation, or how they would react were they to learn about such behavior. In jury research, this method is standard because, in principle, in the United States researchers are not admitted to the jury box. Even if direct observation is not legally prohibited or technically impossible, researchers may prefer a vignette study since it makes identification easier. In a between-subjects design, participants are randomly assigned to different versions of the vignette. Any treatment effect may then be traced back to the difference between the scenarios. In a within-subjects design, every participant reacts to more than one (qualitatively different) vignette. If the reaction differs, this difference must result from having seen the earlier vignettes.

To illustrate, consider a study on the standard of proof in civil litigation (Zamir and Ritov 2012). In the United States, the official standard is preponderance of the evidence. In the literature, this standard is frequently translated into the posterior probability of the claim being well founded to be above 50%. Earlier evidence suggests that triers actually require a higher probability. The study proceeds in two steps. It first has separate groups of law students rate the persuasiveness of the plaintiff's case in three different scenarios. Another group of law students is asked whether they would find for the plaintiff in each case. Despite the fact that the mean rating of persuasiveness is above 50% for all three scenarios, much less than half of all participants declare that they would find for the plaintiff. The result has been replicated with professional lawyers. In the second step, a new group of students is handed a questionnaire that explores potential explanations of the effect. The one explanation that stands out is what the authors call omission bias. Participants believe that judges wish to avoid responsibility for finding for the plaintiff despite the fact that the evidence is weak. Participants believe that judges feel less responsibility if they erroneously dismiss the claim.

Vignette studies are experiments in which the experimenter manipulates the scenarios. Participants are randomly assigned to different versions, or they see different versions over time. Sometimes participants are also alien to the situation the experimenter wants to explain. Often participants are students, although the study wants to explain the behavior of the general public, or of legal officers. Many researchers see this as a limitation, and prefer giving the vignettes to a sample that is representative for the general public (12 of the 23 vignette studies published in *JELS*) or to legal officers (six papers), usually citizens on jury duty.



## 2.5 Lab Experiment

The standard tool of experimental economists is the lab experiment. There are seven main differences from a vignette study. (1) The main, if not the exclusive, dependent variable is a choice. This choice is incentivized. What participants decide, and how they interpret the situation in preparation for their decision, directly matters for the payoff they receive. (2) The researcher's interest lies in abstract effects. To test for a hypothesized effect, the design is free from context. The typical design is a game. The benchmark solution is provided by the response of a person exclusively motivated by pecuniary gain, and in possession of unlimited cognitive abilities. Usually, this prediction is contrasted with an alternative hypothesis based on a richer utility function, or assuming less than perfect cognition. (3) Most economic experiments are interactive. Participants are not studied in isolation but as they interact. (4) Many economic experiments repeat a stage game multiple times. This is done in the interest of studying how effects unfold over time. (5) There is a culture in economic labs that forbids deception. This rule is meant to improve identification. Experimenters want to be sure that any effects indeed result from their manipulation, not from participants second-guessing how the experimenter is trying to trick them. (6) Economic experiments are usually completely computerized. They are run in a computer lab. Complete anonymity is guaranteed. Usually all communication is through choices. These precautions also aim at better identification. (7) Often hypotheses are derived from a formal model. Actually, the experimental methodology is meant to directly map formal economic theory, and game theory in particular (more discussion of the different experimental paradigms in Hertwig and Ortmann 2001).

Compared with the remaining empirical methods, lab experiments put most stress on internal validity. Predictions are as precise as possible, and as clearly grounded in explicit theory as possible. Observations are made as credible as possible. The design tries hard to exclude alternative explanations for the treatment effect, even those resulting from the construction of the situation. The price economic experiments pay for this rigor is less external validity. The policy problem that motivates the experiment is not directly visible. Economic experiments make a contribution to the policy discourse by isolating one effect. There is always reason to discuss whether alternative explanations or additional factors still support intervention. Yet by stripping the situation from all context, and by translating it into a naked incentive structure, one is able to study this driving force.

I illustrate the potential and the limitations of economic experiments for legal issues with one of my own papers (Engel and Kurschilgen 2011). German copyright law has a seemingly odd provision. If a work, say a film, turns out to be a blockbuster, those who have contributed to it may sue the producer and claim additional remuneration. We have translated this provision into a sequential two-person game. At the beginning of the game, both players receive an equal endowment. Additionally, one player holds an unlabeled commodity. The other player can offer to buy this commodity. It is known that the commodity either has little value or is very precious. All gains from trade lie

with the buyer. If the seller accepts, the deal is struck. Otherwise this round ends and both players keep their endowments. If the commodity is traded, Nature determines the value of the commodity. In the final stage, both players can impose harm on their counterpart, at a price to themselves. In the treatment, three stages are added. In the first stage, after Nature has decided, a third player determines “the appropriate purchase price.” Her decision is kept confidential. In the next two new stages, the original players have a chance to renegotiate, using the same protocol as before. If negotiations fail, the third party’s decision becomes effective.

We hypothesized that the rule standing in for the German copyright provision would affect how the parties judged the fairness of the deal, both *ex ante* and *ex post*, and that this would affect their choices. Results support this prediction. If the rule is in place, buyers offer lower prices, and more deals are struck. This is why the rule turns out to be efficient. In German legal discourse, the rule is mainly justified by its purported effect on *ex post* fairness, though. We qualify this expectation. Third parties indeed almost exclusively split gains from trade equally, despite the fact that the buyer carried all the risk and thereby insured the seller. Yet sellers themselves did not see unfairness. Hardly any seller used the punishment option. In a result unpredicted by the German legal debate, buyers were much more likely to exhibit *ex post* discontent. Their willingness to punish the seller was, however, reduced by the rule.

## 2.6 Alternative Empirical Methods

My presentation of empirical methods for behavioral legal analysis has mirrored the methods that have actually been used in the publications in *JELS*. One prominent methodological alternative has not yet made it into this journal, the field experiment. In a field experiment, the experimenter directly intervenes in the social phenomenon she wants to understand. This method is promising in that it combines high external with high internal validity. There are two main challenges. The first challenge is practical. It is technically often not easy, politically often problematic, and maybe just plain illegal if some individuals are randomly deprived of treatment the researcher herself expects to be efficient and beneficial. The second challenge is methodological. However hard the researcher tries, precisely because intervention is in the field, there is less experimental control. The validity of randomization hinges on the definition of the pool from which participants are drawn. These participants cannot be completely standardized, so that treatment effects may result from alternative causes for which the researcher must try to control. Finally, it is standard in field experiments not to reveal manipulation. This may lead to ethical concerns.

The following is an illustration how this method can be applied to a legal issue (Listokin 2010). Used iPods were auctioned off on eBay with randomly varied return policies. Some iPods came with a satisfaction guaranteed policy, others with an explicit warranty that resembled the default warranty of the Uniform Commercial Code, and still other iPods were sold “as is.” Finally, a batch of iPods was silent regarding the

return policy. The mean price paid in the auction shows that consumers are sensitive to the warranty. Prices were highest if buyers were free to give the iPod back. Prices were lowest if a guarantee was expressly excluded. In the two remaining conditions, prices were in the middle, and not statistically distinguishable. The author concludes that buyers assume a safety level as guaranteed by the majoritarian default of the code if the contract is silent on the warranty.

Economists sometimes use simulation to show on which sets of parameters a problem is conditioned. Simulation presupposes the complete definition of a mechanism. In each simulation run, one parameter is changed. Simulation is also useful if one expects one or more processes to be random, with a defined nature of the disturbance. Recently, simulation has also been used in the behavioral legal literature. In the very differentiated empirical literature on lineups, it is held to be established that eyewitnesses are more reliable if they base their recognition judgment on an absolute, rather than a relative, criterion. It is argued that eyewitnesses might accept the relatively closest analogue to their recollection, rather than the one person who is so similar to memory traces that there is no room for doubt. The paper translates this claim into a formal model of recognition and manipulates memory accuracy, the degree of similarity between perpetrator and foil, and the decision criterion. It turns out that trying to meet an absolute threshold does not minimize false positives under all circumstances. A relative approach performs better if the witness's memory is relatively accurate and an innocent suspect is fairly similar to the perpetrator (Clark, Erickson, and Breneman 2011).

All of the foregoing methods are quantitative. They are meant to generate evidence the relevance of which is judged by way of frequentist statistics. Occasionally, law and psychology researchers instead use qualitative methods. One recent study interviewed defendants who just had pled guilty, and explored in which ways they had understood the plea inquiry. It turned out that errors were widespread. Two-thirds of the sample were correct on less than 60% of the questions the researchers asked them (Redlich and Summers 2011). The main advantage of qualitative empirical research results from the fact that it is not bound by a set of dependent variables that have been defined *ex ante*. The researcher can give each individual observation full justice, and can use the evidence to learn about hitherto neglected aspects of the issue at hand. The main drawback directly results from this advantage. Since the dependent variable is not standardized, it is more difficult to assess whether individual observations generalize.

### 3 FUTURE DIRECTIONS

---

Outside criminology and law and psychology, the empirical legal movement is still young. This also holds for its behavioral branch. In conclusion I sketch potential paths for future development.

When presenting the most prominent empirical methods in greater detail, it has become clear that there is a trade-off between external and internal validity. A straightforward reaction is combining more than one empirical method on the same research question. Occasionally, this is even done within the same publication. A study on racial bias in bankruptcy uses this approach (Braucher, Cohen, and Lawless 2012). In the United States, consumers can file bankruptcy under chapter 7 or under chapter 13. While the latter procedure may be advantageous if the debtor wants to protect valuable assets, consumers usually use the former procedure since it is less onerous and less costly. The study has two parts. In the first part it shows that African Americans are disproportionately more likely to file under chapter 13, even after controlling for relevant sociodemographic factors. A vignette study randomly asks bankruptcy lawyers who usually represent consumers to give advice to a couple with Christian names that suggest an Afro-American or a Caucasian background. Bankruptcy lawyers advise the former couple significantly more frequently to file under chapter 13.

Quantitative studies quantify the trust one may have in the result by a significance test. Nonetheless, researchers may have overlooked a qualifying factor. Despite their attempts at securing randomness, experimentalists may have worked with an atypical sample. Inadvertently, a feature of the design of an experiment that seemed innocent may have been critical. Ultimately, the law should therefore be hesitant to derive normative conclusions from a single empirical study. A procedure that is standard in medicine is still very rare in law, the replication of findings (for an exception, Hall 2010). Equally rare is the reanalysis of empirical data with alternative statistical models (for an exception see Goodsell, Gronlund, and Carlson 2010). Finally meta-analysis, that is, the structured, quantitative analysis of findings from a whole line of research, is thus far confined to law and psychology, and almost exclusively to forensic psychology (e.g., Steblay, Dysart, and Wells 2011). As the field matures, all of these methods for assessing the robustness of findings and for better understanding framework conditions should become more prominent.

Economic experiments have been invented as tests for formal economic theory. Decades ago, formal economic theory made headway into law. In many subfields of law, and in private law in particular, economic theorizing is fairly advanced. Thus far, tests of formal law and economics hypotheses are still rare. One example is a paper that first models the effect of split-award statutes on negotiations between tortfeasors and victims (Landeo, Nikitin, and Babcock 2007). Under these statutes, the plaintiff receives only a portion of punitive damages, while the remainder is paid to the state. The authors translate the situation into a sequential game, solve for equilibrium, and test the resulting predictions in the lab. They find that these statutes do not affect the level of care, but reduce the likelihood of trial, and the total litigation cost born by the community of parties.

Disciplines have their traditions. Traditions may result from historical contingency. Yet at least in the long run, traditions are likely to converge to the functional needs of a discipline. Arguably the different experimental traditions in psychology and economics reflect differences in the dominant research questions. The same claim seems plausible for the different approaches to the analysis of field data in criminology

and econometrics. These observations suggest that, in the long run, behavioral legal researchers might want to develop their own, discipline-specific empirical methods. Three challenges are likely to be particularly pronounced if an attempt is made to introduce such evidence into legal argument.

Lawyers frequently want to judge proposals for institutional intervention. This is obvious if a legal scholar makes a contribution to the legal policy discourse. Similar questions are asked by doctrinal lawyers if they have to decide on the constitutionality of institutional reform, or if they have to rely on teleological interpretation to resolve an ambiguity of the text. Often the purported effect of the intervention hinges on assumptions about the behavior of typical addressees. Yet the policy question is only partly answered by a list of relevant findings from basic behavioral research. It ultimately matters whether the specific intervention delivers on its promises, without having too many undesirable side-effects. Answering this question might require testing entire institutions, rather than isolated effects. Of course, if one does, one partly loses control since institutions are lumpy responses to lumpy perceived problems. Nonetheless, the knowledge generated that way may be more valuable since this very combination of effects is likely to be at work in legal practice. Related to this, legal institutions are very rarely designed from scratch. The typical situation is institutional reform. The legislator intervenes in the interest of improving what it thinks fell short of normative expectations. Therefore often the critical behavioral question is how addressees will react to an intervention meant to induce behavioral change. In an experiment, this can be reflected by a sequential design that focuses on the difference before and after the introduction of the tested legal institution.

Ultimately legal rules are meant to decide disputes. Negotiators, administrators, and courts must settle disputes in good time. To do this effectively, they must reflect what the parties see as essential features of the case. Both the characteristic time pressure and the typical level of specificity do not easily go hand in hand with empirical methods that have been developed in the social sciences to answer questions of basic science. Nonbehavioral areas of law have found solutions for the resulting methodological challenges. Merger simulations provide a good illustration (Budzinski and Ruhmer 2010). They rely on formal economic theory of industrial organization. They capitalize on rigorous econometric work in this subdiscipline of economics. But the actual simulation does not try to build all the methodological safeguards into the simulation model. And since no closed form solution is required, simulation may simultaneously address many dimensions of a merger case, even if there is debate over how to model or measure them. For appropriate legal applications, behavioral researchers might develop similar tools.

Essentially, all practical legal problems are ill defined, but procedurally contained. By contrast, typically the empirical methods developed in economics and psychology address well-defined problems in an artificial setting. Usually, researchers do not aim at mapping the design of their study to the procedural framework that contains the legal reaction in the field. Ill-defined problems do not have one normative solution. This is one of the reasons why judges and juries cannot be replaced by law-making machines. The perception of the case tends to have an idiosyncratic element. Courts decide upon the lives of real people. Anything these persons offer as a relevant

feature of the situation has in principle to be taken into consideration—but only if it is introduced into court procedure the proper way. Once the procedure is closed, fresh argument is precluded, except if there is a right to appeal, to list only two of the many features of the procedural framework. Quasi experiments might be one way of casting light on the interaction between an incomplete set of facts and an institutional setting for processing them. The technique has been developed by Nobel Prize winner Reinhard Selten, but has been rarely used. He was interested in studying investment behavior of firms in a complex setting. To that end he assigned fixed roles to student participants, and had them interact over an entire term, with the goal being, for each group of participants, to jointly develop what they believed to be the optimal investment policy (Selten, Mitzkewitz, and Uhlich 1997). By the same token, participants could be invited to develop a litigation strategy in a case assigned to them by the experimenter, and reacting to the idiosyncratic ways in which their case unfolds in their randomly composed group of parties, judge, jury members, and maybe advisors, witnesses, and experts.

In many respects, empirical behavioral research on legal issues still is a nascent endeavor. Inevitably, this chapter has only been able to provide a snapshot of a rapidly moving field. There are two take-home messages though. The field is highly differentiated and capitalizes on multiple methods from many neighboring disciplines. There is considerable room for improvement. Behavioral legal researchers should take methodological standards that have developed in the more mature neighboring fields more seriously. And they should spend more energy on developing empirical methods that directly map the needs of their own discipline.

## ACKNOWLEDGMENTS

---

Helpful comments by Angela Dorrough and Pascal Langenbach are gratefully acknowledged.

## REFERENCES

---

- Blundell, Richard, and Monica Costa Dias (2009). "Alternative Approaches to Evaluation in Empirical Microeconomics." *Journal of Human Resources* 44(3): 565–640.
- Braucher, Jean, Dov Cohen, and Robert M. Lawless (2012). "Race, Attorney Influence, and Bankruptcy Chapter Choice." *Journal of Empirical Legal Studies* 9(3): 393–429.
- Brinig, Margaret F., and Nicole Stelle Garnett (2012). "Catholic Schools and Broken Windows." *Journal of Empirical Legal Studies* 9(2): 347–67.
- Buccafusco, Christopher, and Christopher Sprigman (2011). "Valuing Intellectual Property: An Experiment." *Cornell Law Review* 96: 1–46.
- Budzinski, Oliver, and Isabel Ruhmer (2010). "Merger Simulation in Competition Policy. A Survey." *Journal of Competition Law and Economics* 6(2): 277–319.



- Cardi, W. Jonathan, Randall D. Penfield, and Albert H. Yoon (2012). "Does Tort Law Deter Individuals? A Behavioral Science Study." *Journal of Empirical Legal Studies* 9: 567–603.
- Clark, Steven E., Michael A. Erickson, and Jesse Breneman (2011). "Probative Value of Absolute and Relative Judgements in Eyewitness Identification." *Law and Human Behavior* 35(5): 364–80.
- Eisenberg, Theodore, and Christoph Engel (forthcoming, 2014). "Assuring Civil Damages Adequately Deter. A Public Good Experiment." *Journal of Empirical Legal Studies*.
- Engel, Christoph (2005). *Generating Predictability: Institutional Analysis and Institutional Design*. Cambridge: Cambridge University Press.
- Engel, Christoph (2010). "The Behaviour of Corporate Actors. A Survey of the Empirical Literature." *Journal of Institutional Economics* 6: 445–75.
- Engel, Christoph, and Michael Kurschilgen (2011). "Fairness Ex Ante and Ex Post: Experimentally Testing Ex Post Judicial Intervention into Blockbuster Deals." *Journal of Empirical Legal Studies* 8: 682–708.
- Frakes, Michael (2012). "Defensive Medicine and Obstetric Practices." *Journal of Empirical Legal Studies* 9(3): 457–81.
- Gazal-Ayal, Oren, and Raanan Sulitzeanu-Kenan (2010). "Let My People Go: Ethnic In-Group Bias in Judicial Decisions. Evidence from a Randomized Natural Experiment." *Journal of Empirical Legal Studies* 7(3): 403–28.
- Goodsell, Charles A., Scott D. Gronlund, and Curt A. Carlson (2010). "Exploring the Sequential Lineup Advantage Using Witness." *Law and Human Behavior* 34(6): 445–59.
- Hall, Matthew (2010). "Randomness Reconsidered. Modeling Random Judicial Assignment in the US Courts of Appeals." *Journal of Empirical Legal Studies* 7(3): 574–89.
- Hertwig, Ralph, and Andreas Ortmann (2001). "Experimental Practices in Economics. A Methodological Challenge for Psychologists?" *Behavioral and Brain Sciences* 24(3): 383–403.
- Kahneman, Daniel, and Amos Tversky (2000). *Choices, Values, and Frames*. Cambridge: Russell Sage Foundation; Cambridge University Press.
- Klick, Jonathan, and Thomas Stratmann (2007). "Diabetes Treatments and Moral Hazard." *Journal of Law and Economics* 50(3): 519–38.
- Landeo, Claudia M., Maxim Nikitin, and Linda Babcock (2007). "Split Awards and Disputes: An Experimental Study of a Strategic Model of Litigation." *Journal of Economic Behavior & Organization* 63(3): 553–72.
- Leamer, Edward E. (1983). "Let's Take the Con out of Econometrics." *American Economic Review* 23: 31–43.
- Listokin, Yair (2010). "The Meaning of Contractual Silence: A Field Experiment." *Journal of Legal Analysis* 2(2): 397–416.
- Posner, Richard A. (2011). *Economic Analysis of Law*. New York: Aspen Publishers.
- Rachlinski, Jeffrey J. (2011). "The Psychological Foundations of Behavioral Law and Economics." *University of Illinois Law Review* 2011: 1676–96.
- Redlich, Allison D., and Alicia Summers (2011). "Voluntary, Knowing, and Intelligent Pleas: Understanding the Plea Inquiry." *Psychology, Public Policy, and Law* 18: 626–43.
- Redlich, Allison D., Alicia Summers, and Steven Hoover (2010). "Self-Reported False Confessions and False Guilty Pleas among Offenders with Mental Illness." *Law and Human Behavior* 34(1): 79–90.
- Rose, Mary R., Shari Seidman Diamond, and Kimberly M. Baker (2010). "Goffman on the Jury." *Law and Human Behavior* 34(4): 310–23.



- Schwartz, Gary (1994). "Reality in the Economics of Tort Law: Does Tort Law Really Deter?" *UCLA Law Review* 42: 377–444.
- Selten, Reinhard, Michael Mitzkewitz, and Gerald R. Uhlich (1997). "Duopoly Strategies Programmed by Experienced Players." *Econometrica* 65: 517–55.
- Simon, Dan, Daniel C. Krawczyk, and Keith J. Holyoak (2004). "Construction of Preferences by Constraint Satisfaction." *Psychological Science* 15: 331–36.
- Stebly, Nancy K., Jennifer E. Dysart, and Gary L. Wells (2011). "Seventy-Two Tests of the Sequential Lineup Superiority Effect. A Meta-Analysis and Policy Discussion." *Psychology, Public Policy, and Law* 17(1): 99–139.
- Sunstein, Cass R., Ed. (2000). *Behavioral Law and Economics*. Cambridge: Cambridge University Press.
- Tor, Avishalom (2008). "The Methodology of the Behavioral Analysis of Law." *Haiifa Law Review* 4: 237–327.
- Yang, Y. Tony, David M. Studdert, S. V. Subramanian, and Michelle M. Mello (2012). "Does Tort Law Improve the Health of Newborns, or Miscarry? A Longitudinal Analysis of the Effect of Liability Pressure on Birth Outcomes." *Journal of Empirical Legal Studies* 9(2): 217–45.
- Zamir, Eyal, and Barak Medina (2011). *Law, Economics, and Morality*. Oxford: Oxford University Press.
- Zamir, Eyal, and Ilana Ritov (2012). "Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation." *Journal of Legal Studies* 41(1): 165–207.
- Zeiler, Kathryn, and Charles R. Plott (2005). "The Willingness to Pay / Willingness to Accept Gap, the Endowment Effect, Subject Misconceptions and Experimental Procedures for Eliciting Valuations." *American Economic Review* 95: 530–45.

## CHAPTER 6

---

# BIASING, DEBIASING, AND THE LAW

---

DANIEL PI, FRANCESCO PARISI, AND BARBARA LUPPI

HISTORICALLY, economists understood rational choice theory as both a normative and a descriptive model of human behavior; it was thought to *prescribe* how people ought to behave, and also to *describe* how they do in fact behave. However, in the 1950s, a handful of economists and psychologists began to question the latter claim—that rational choice theory is able to predict/explain real-world behavior (Allais 1953; Simon 1955). This skeptical line of inquiry gained increasing momentum in the 1970s (Tversky and Kahneman [1974, 1981]; Kahneman and Tversky 1979), and by the early twenty-first century behavioral economics has become one of the main branches of economic research.

Rational choice theory is grounded on the simple premise that economic actors are (private) welfare maximizers. In a nutshell, the idea is that when presented with choices, the rational actor looks to the payoffs entailed by his available options, and selects the choice that yields the greatest expected utility. The assumption that individuals seek to advance their interests by making choices that maximize their payoffs seems tautological, and yet a wealth of experimental and empirical data seem to suggest that this proposition often fails to obtain in reality.

Of course, classical economists never assumed that all individuals optimize at all times; rather, the prevailing view was that particular idiosyncratic deviations from the predictions of rational choice theory would cancel out over time, such that anomalous departures would lead to variance, but not to changes in the average behavior of a population. To the extent that laws must be general and addressed to the population at large, designing laws that would work well for the representative (i.e., average) member of the population was ultimately viewed as the best that lawmakers could do, leaving the value of rational choice theory undiminished for practical application in the analysis of law. However, the experimental and empirical data do not merely suggest the presence of outliers—a banal observation, surely—nor mere variance around estimated means, but rather more profoundly that people *systematically* diverge from the predictions of rational choice theory.

These findings present serious obstacles for economic theory and the social policies motivated by it. Law and economics scholars have conventionally regarded

social welfare to be an aggregation of the private welfare of individuals. Thus, when actors fail to maximize their private welfare, the calculation of incentives and social welfare, which relied upon the rational actor hypothesis, may be radically disturbed.

In this chapter, we will present an overview on the use of legal mechanisms to combat inefficiencies arising from biased judgment and behavior. We will also explore how heuristics and biases may be exploited to foster efficiency in the presence of incentive misalignment (e.g., market failures, collective action problems, hyperbolic discounting and time inconsistency, principal-agent problems). To that end, we introduce two new categories of social engineering under bounded rationality: two hitherto unnoticed logical counterparts to debiasing and insulating strategies, which we call “benevolent biasing,” and “cognitive leveraging.”

We will begin in section 1 with a brief exposition of background theory and terminology. In section 2, we will consider debiasing and insulating strategies as correctives when biases result in departures from efficiency under perfect competition. In section 3, we will consider benevolent biasing and cognitive leveraging strategies as legal instruments to correct other forms of inefficiency. We conclude in section 4 with a summary of the foregoing material.

## 1 BACKGROUND AND TERMINOLOGY

---

The jargon of behavioral law and economics can be the source of much confusion and misunderstanding, so we will begin this section with some brief definitions. For the sake of brevity, precision, and clarity, we will present formal descriptions where they may be helpful. We have endeavored to avoid inventing new terms, as well as assigning controversial meanings to existing terms, though in some cases we offer definitions that are more precise than past usage. We then give some caveats to our theoretical approach, followed by examples of heuristics and biases from the literature.

### 1.1 Definitions

We begin by defining a generalized private payoff function as an objective baseline. Let  $P_i(c)$  denote the objective payoff of individual  $i$  choosing  $c$ . The optimal choice for individual  $i$ , given a domain of possible choices  $D$  will be

$$x_i^* = \max_{x \in D} P_i(x).$$

Now let us contrast this with  $i$ 's perceived (i.e., subjective) payoffs, given a perceived domain of alternatives  $D^o$ , which we will denote as  $P_i^o(c)$ . We may now more precisely define several key terms in describing nonrational behavior.

First, let us consider the case where  $i$  suffers from no biases—that is, where  $i$  is perfectly rational. When  $i$ 's subjective payoff is equal to his objective payoff,<sup>1</sup> and  $i$  chooses  $x^*$ , we shall refer to such an individual  $i$  as an *unboundedly rational* actor.

Second, let us consider the case where  $i$  optimizes on his subjective payoff function, which diverges from his objective payoff.<sup>2</sup> Let us call his subjective optimum  $x^o$ , such that

$$x_i^o = \max_{x \in D^o} P_i^o(x).$$

and let us refer to such an individual  $i$  as a *boundedly rational* actor. When such an individual makes a decision, he *believes* he is maximizing his payoff, but because he is mistaken about the benefits, costs, or domain of his available choices, he behaves suboptimally.<sup>3</sup>

Third, when individuals make choices using a decision-making procedure other than optimization, we refer to the choice as *nonrational*. That is, when the mechanism for making a choice is not maximization, then the decision is nonrational.

Fourth, when individuals make suboptimal choices systematically—for significant numbers of people (or regularly over time for an individual)—let us refer to the decision-makers and their decisions as *biased*. If the deviation of  $x_i^o$  from  $x_i^*$  exhibits some recognizable characteristic, we may identify the bias as belonging to a particular type (e.g., overoptimism bias, hindsight bias, availability bias).

It now bears remarking upon methodological confusions that may arise from the foregoing definitions. Psychologists may object to our description of biases as “systematically suboptimal behavior.” Insofar as our definition fails to do justice to the multifaceted complexities of the human mind and the processes of decision-making, we concur wholeheartedly. Nevertheless, to the extent that “bias” may mean more than merely “systematically suboptimal behavior” to a psychologist, for the purposes of applied economics, much of the richer meaning seems to us irrelevant. This is more acutely the case when discussing “debiasing” and “biasing” strategies, which seek to align irrational decision-making with rational choice, where the relevant questions are (1) to what extent is the biased behavior suboptimal, and (2) what policies will cause biased individuals to make optimal decisions (while holding incentives and preferences fixed).

We should further articulate some elementary distinctions and caveats for the sake of clarity. Debiasing (and biasing) should not be confused with altering preferences for the simple reason that a preference (however deviant it may be) is not a “bias.” For example, drugs designed to eliminate the pleasure that consumers derive from cigarettes or alcohol (with the aim of helping them quit) should not be regarded

<sup>1</sup> I.e., where  $(\forall x \in D)(x \in D^o \wedge P_i(x) = P_i^o(x))$ .

<sup>2</sup> I.e., where  $\neg(\forall x \in D)(x \in D^o \wedge P_i(x) = P_i^o(x))$

<sup>3</sup> This is not strictly true. It is possible that in some cases  $\neg(\forall x \in D)(x \in D^o \wedge P_i(x) = P_i^o(x))$  and  $x^* = x^o$ . That is, boundedly rational individuals can sometimes make optimal choices. For

as debiasing.<sup>4</sup> Debiasing does not change individuals' preferences. It eliminates errors in judgment. This is not to say that altering preferences may not be a useful or desirable end—indeed this may often be an important policy goal—however, changing preferences and debiasing are entirely independent phenomena, and they should not be conflated.<sup>5</sup>

Consider the widespread problem of obesity. There may be two causes of overeating. First, people may misjudge the facts: they may underestimate the negative health consequences of obesity; or they may overestimate what a healthy weight is; or they may underestimate the caloric value of the foods they eat. Second, people may simply value the pleasure they get from eating more than the detrimental health effects of overeating. The first category of causes may be corrected by debiasing; however the second category is not the result of bias. Given especially hedonistic preferences (or hyperbolic time-discounting), an unboundedly rational individual may continue to overeat, simply because the pleasure he derives from eating unhealthy foods outweighs the increased probability of health problems.

On this point, some social scientists will distinguish between “bounded rationality” and “bounded willpower” (and also “bounded self-interest”). While these distinctions may be analytically convenient in certain contexts, we do not think, for our purposes, they are relevant.<sup>6</sup> Bounded willpower and bounded self-interest may be modeled in terms of preferences; and any objection to this will not hinge on the choice of formal method so much as its interpretation.

It may now be complained that we have adopted what Elster (1985) disparagingly called a “thin” theory of rationality, that is, a theory that regards preferences as exogenous and fixed. Unfortunately, a thorough discussion of the philosophical foundations of rationality would take us far afield of the present topic, and we should not like to do injustice to what is a profoundly important issue by giving it a cursory treatment. Instead, we will remain agnostic on the point, while observing that the foregoing framework does not preclude one from adopting a “thick” theory of rationality—unless one were committed to the semantic (not substantive) claim that “debiasing”

example, when a misperception in costs is entirely offset by complementary misperceptions in benefits—or more trivially, when the domain of available choices contains only one member.

<sup>4</sup> The drug Chantix operates by blocking neuroreceptors that cause smokers to feel pleasure from consuming nicotine; while the drug Antabuse similarly interferes with the pleasure that drinkers feel from consuming alcohol.

<sup>5</sup> Bilz and Nadler, chapter 10 in this volume, offer an excellent discussion about the law's capacity to change attitudes (and the desirability of doing so). We wish to be clear that by articulating that such policies, designed to produce a change in preferences, do not count as “debiasing,” we are not claiming that they are undesirable or unworthy of analysis (quite to the contrary). Rather, we simply wish to highlight the fact that it does not fall under the definition of “debiasing,” and in a relatively young field, afflicted with a rapidly ballooning and often inconsistently applied nomenclature, getting clear on the meanings of terms is likely to forestall considerable future confusion and controversy.

<sup>6</sup> Additionally, we find the coining of buzzwords for newly recognized phenomena to be a troubling trend, insofar as they discourage generalizations to underlying principles and encourage the production of laundry lists of ever-more-thinly sliced bare observations claiming to be “theories.”

includes changing preferences. So long as one agrees on the meaning of the nomenclature, it will not matter whether one adopts a “thick” or “thin” conception of rationality with respect to debiasing.

Also, debiasing should not be confused for incentive alignment. Altering incentives to compensate for biased behavior should not be regarded as either an insulating or debiasing strategy. Biases change how individuals respond to incentives. Undoubtedly, changing incentives will have *similar effects* to debiasing in some cases; however, it would be a mistake to conflate debiasing with incentive alignment. Incentive alignment creates incentives for individuals to make socially optimal choices. Debiasing does not change those incentives, but rather helps individuals to see them more clearly (or in the case of insulating strategies, restricts the domain of possible choices).

## 1.2 Examples

Let us now consider some examples of common biases identified in the literature of behavioral economics. One of the most easily identifiable biases is “overoptimism,” which affects individuals’ judgments of probabilities.<sup>7</sup> When individuals are faced with uncertain situations, they tend to overestimate probabilities that good things will happen to them and underestimate probabilities that bad things will happen to them.

Overoptimism takes several more specific forms. For example, the “better than average” effect describes the all-too-common fact that many people consider themselves better than average with respect to some skill or ability.<sup>8</sup> Svenson (1981) found that a majority of survey respondents considered themselves “better than average” automobile drivers. No doubt some drivers must be better than average, but obviously, it cannot be the case that *more* than 50% of drivers are *better* than the 50th percentile driver. The better-than-average effect may lead individuals to underestimate the probability of punishment in a criminal law context, overestimate the effectiveness of their precautions in tort law context, underestimate the probability of breaching in a contract context—indeed, the types of cases where better-than average-effect may distort the subjective evaluation of probabilities are too many to enumerate.

Another species of overoptimism is the “blind spot bias,” which is an interesting second-order incarnation of better-than-average effect (Pronin, Lin, and Ross 2002). Individuals suffering from blind spot bias may be well informed about biases,

<sup>7</sup> The seminal paper on overoptimism is Weinstein (1980), though elements of overoptimism were identified as early as Lund (1925) and Cantril (1938). Wengert and Rosén (2000) offers a good overview of subsequent research. Williams, chapter 13 in this volume, provides an up-to-date overview of overoptimism, as well as ambiguity aversion and the certainty effect.

<sup>8</sup> Alicke and Govorun (2005) provide a complete review of “better than average” effect, which is also sometimes called “above average effect,” or the “Lake Wobegon effect” (after the public radio program *A Prairie Home Companion*, set in the fictional town of Lake Wobegon, where “all the women are strong, all the men are good looking, and all the children are above average.” Lee (1991).

but underestimate their own susceptibility to them (i.e., they are biased about being biased). That is, even when they are informed about commonly occurring biases, they may fail to correct such judgment errors in their own decision-making, because they optimistically regard themselves as immune to or less influenced by destructive biases; they have a “blind spot” with respect to their own decision-making and judgment-forming processes. Blind spot bias is of particular concern when debiasing, because individuals suffering from blind spot bias will tend to be especially resistant to many debiasing remedies (e.g., education is unlikely to succeed when biased individuals do not believe the improved information is applicable to themselves).

Overoptimism may be partially explained by “representativeness bias.” Representativeness bias is the result of a heuristic whereby individuals generalize from particular “representative” cases, even when better general statistical information exists (Tversky and Kahneman 1974). It is easy to see why the representativeness heuristic might be useful in everyday reasoning—specific information is often more accurate or more useful than general information, and favoring it may represent a useful cognitive shortcut when assessing probabilities. However, when this assumption fails, the representativeness heuristic can lead to bizarre and erroneous judgments. Overoptimism may result from representativeness when, for example, individuals infer from knowledge about a specific car accident involving a bad driver that bad driving correlates with accidents more highly than it does in fact. In this case, even if they believe they are average drivers, they will systematically underestimate the probability that they will be involved in an accident, because the representativeness heuristic will have generated an underestimation of the probability that average drivers are involved in accidents.

Another common cognitive illusion is “hindsight bias,” which describes inconsistent probability estimates with respect to an individual’s temporal position relative to some event (Fischhoff 2003).<sup>9</sup> When looking back at the probability of an event occurring, which did in fact occur, individuals tend to overstate the probability of its having occurred. That is, if the probability of event  $E$  occurring is  $p(E)$ , then an individual suffering from hindsight bias will estimate the probability as  $q(E)$ , such that  $q(E) > p(E)$ , even though the same individual would have given an accurate estimate of the event *ex ante*. Knowing that an event *did* happen inflates individuals’ estimates of how likely it was that it *would* happen.

“Framing effects” are a general category of biases, which result from reference point-based estimates of probability.<sup>10</sup> Framing effects arise due to the manner in which a choice is presented. For example, if the cost of registering for some program is \$50 with a \$50 late fee, individuals may be highly motivated to avoid the penalty. However, if the program were offered for \$100 with a \$50 early registration discount, then individuals are observed to be less motivated to register in a timely fashion. Even though the material effect is exactly identical (as between \$100 with a \$50 discount for early

<sup>9</sup> For a history of the development of the theory underlying hindsight bias, see Fischhoff (2007) and Teichman, chapter 14 of this volume.

<sup>10</sup> Kahneman and Tversky’s (1979) “prospect theory” is easily the most prominent example of reference-point based alternative to expected utility theory. Zamir, chapter 11 of this volume, provides a comprehensive overview of loss-aversion.



registration versus \$50 with a \$50 penalty for late registration, *ceteris paribus*), individuals respond differently depending on how the incentive is framed. Notable species of framing effects are connected to the anchoring bias and the endowment effect.

The endowment effect will come up later in our discussion, so it bears saying a word about it now.<sup>11</sup> The endowment effect is motivated by a sort of loss aversion; individuals are observed to systematically value an item that they have more than they would pay to acquire the very same item if they did not have it. That is, an individual may refuse to part with an item he owns for anything less than \$10, but would only pay \$5 to acquire it if he didn't already own it. It is ambiguous whether such an individual values the item at \$5 or \$10,<sup>12</sup> since he seems to reveal different preferences depending on how a potential exchange is framed.<sup>13</sup>

## 2 DEBIASING AND INSULATING STRATEGIES

Let us assume a market in a state of "perfect competition." That is, the aggregation of privately optimal choices leads to the maximization of social welfare.<sup>14</sup> With no loss of generality, let us model this circumstance as a two-person world, and let us call its inhabitants individual 1 and 2. Optimal social welfare would therefore be

$$S^{**} = P_1(X^{**}) + P_2(y^{**}) = P_1(x^*) + P_2(y^*).$$

So long as individuals 1 and 2 make choices that yield the highest private payoffs, the aggregate payoffs for individuals 1 and 2 will also be maximal. Now, if the two individuals are unboundedly rational, then the payoff from  $S^{**}$  will trivially represent an efficient equilibrium, and the law has no useful role to play.

However, if either or both of the two individuals are boundedly rational, then they may make suboptimal choices, and social welfare will therefore also be suboptimal. Suppose that individual  $i$  is boundedly rational and call  $\beta_i$  the "cost" of  $i$ 's bias. He will therefore effect a reduction in social welfare equal to the difference:

<sup>11</sup> Korobkin, chapter 12 of this volume, offers a more comprehensive investigation of this topic.

<sup>12</sup> Loss-aversion explanations will tend to identify the lower valuation as the "true" valuation, and the higher one as biased (Tversky and Kahneman 1991).

<sup>13</sup> Kahneman and Tversky (1979), Tversky and Kahneman (1991), Kahneman, Knetsch, and Thaler (1991), and Thaler (1980, 1985) explained the endowment effect as a manifestation of loss-aversion; however, the endowment effect need not be regarded as a type of loss-aversion. Others have argued that it is the result of changing preferences (e.g., Morewedge et al. 2009 claims that ownership of a good changes the owner's valuation of the good). To the extent that the endowment effect is the result of changing preferences, it is not a bias, and mitigating or eliminating the endowment effect would neither be "debiasing" nor clearly desirable.

<sup>14</sup> We assume Kaldor-Hicks aggregation.

$$\beta_i(x^o) = P_i(x^*) - P_i(x^o).$$

The law now has two possible roles to play in effecting greater efficiency: (1) to optimize “around”  $\beta_i$ ; or (2) to effect  $\min_{x^o} \beta_i$ . The former approach (“insulation”) lets the biases fall where they may, removing outcomes from the reach of biased decision-making; the latter approach (“debiasing”) reduces the effect of the bias by eliminating or reducing  $i$ ’s misperception of benefits, costs, or the domain of options.

The distinction between debiasing and insulation was first pointed out by Jolls and Sunstein (2006). Of course, taken literally, both approaches could be called “debiasing,” insofar as they both seek to mitigate or eliminate the effects of biased judgments from outcomes. However, Jolls and Sunstein distinguished *debiasing* as “debiasing through law,” from *insulation* as “debiasing law.” The difference is that debiasing combats the bias itself by educating or informing biased individuals to eliminate or reduce the bias itself, whereas insulation allows biased individuals to remain biased, and instead seeks to remove biased decision-making (often by removing biased decision-makers) from affecting outcomes.<sup>15</sup>

## 2.1 Insulation

Let us begin by analyzing insulation, which is both simpler to understand and simpler to translate into legal rules. First consider a stylized two-person world, where individual 1 is boundedly rational, and individual 2 is unboundedly rational. Fundamentally, insulating simply means restricting the domain of choices available to individual 1. Auspiciously, since individual 2 is unboundedly rational, the law may often be designed in such a way as to make individual 1’s payoff contingent upon individual 2’s decisions.

For example, in the arena of products liability, it may be the case that typical consumers are prone to suffer from unrealistic optimism, misusing products in such a way that they suffer harm. By imposing liability against manufacturers for such harms, the law transfers the burden of preventing such harm from the consumer to the manufacturer. The rational manufacturer will take measures to prevent consumers from misusing their products, reducing the burden of exercising due precautionary care from the (biased) consumer, even if an *unbiased* consumer would have been the more effective precaution-taker.<sup>16</sup>

<sup>15</sup> Terminologically, this may result in some confusion, since the term “debiasing” refers both to the general category and to a specific subcategory. This is lamentable; however, it will ordinarily be clear from the context whether the general or specific term is intended, and we will therefore use the conventional terminology, rather than inventing new terms.

<sup>16</sup> Of course, Shavell (1980) tells us that manufacturers will reduce activity level rather than increasing precautions. However, because the precautions of manufacturers and consumers are not independent variables, Shavell’s results are not on point here.

In the realm of corporate law, Langevoort (2001) observes that “inside directors” (e.g., those who are also employees of the firm) often suffer from overoptimism.<sup>17</sup> An insulating strategy would remove certain board decisions—that is, those where overoptimism may be particularly problematic—from those individuals, shifting decision-making power to “outside directors,” who tend to be less biased.<sup>18</sup> What is characteristic about insulation here is that the law leaves inside directors as biased as it found them. It merely limits their participation in the decision-making process.<sup>19</sup>

Of course, insulation is not costless. For example, it may sometimes be the case that the consumer precautions are more effective than manufacturer precautions. In the specific example, the cost of insulation is ultimately borne by all consumers, meaning that unbiased consumers are paying a higher price to make the product safer for biased consumers. These distributive effects may alter incentives. However, if the consequences of insulation result in social welfare closer to the optimum than the consequences of bias (and secondary effects), then an insulating strategy may be a reasonable second-best result.

Formally, let us call the social welfare maximizing rule  $A$ , and some alternative (insulating) rule  $B$ . Let us call the difference in welfare  $\alpha = S^A(x^*, y^*) - S^B(x^*, y^*)$ . Let us suppose now that one party is susceptible to bias, such that rule  $A$  is affected (i.e.,  $S^A(x^o, y^*) < S^A(x^*, y^*)$ ), but rule  $B$  is not (i.e.,  $S^\uparrow B(x^o, y^{\uparrow*}) = S^\uparrow B(x^{\uparrow*}, y^{\uparrow*})$ ). It will only be the case that Rule  $B$  should now be preferred if  $\alpha < \beta$ . That is, the reduction in welfare from selecting a suboptimal rule—for example, imposing a suboptimal liability rule in a tort, or excluding inside directors from decision-making—should be less than the reduction in welfare that would have been caused by biased decision-making.

Obviously, it will often be the case that *both* parties in an interaction will be biased to some degree—though perhaps in different ways. However, the foregoing analysis remains valid. If  $\beta_1 > \beta_2 > 0$ , then an insulating strategy will reduce the domain of individual 1’s available choices, possibly by placing the burden on individual 2. Even though individual 2 may also be biased, because he is *less* biased than individual 1, insulating will at least spare us from the greater of the two biases. However, when both parties are prone to particularly destructive biases, it will clearly be less likely that such an insulating strategy would be effective, since the threshold would then be  $\alpha + \beta_2 < \beta_1$  (where  $\beta_2$  represents the cost of the less destructive bias).

Next consider a two-person world, where both parties are biased, and both parties’ biases are equally destructive (i.e.,  $\beta_1 \approx \beta_2$ ). In this case, insulating strategies that shift decision-making to less biased parties will obviously be ineffective, since both parties’ biases would be equally destructive. To be effective, insulating strategies would have to reduce the

<sup>17</sup> Langevoort (2001) suggests that the personality traits that lead to an employee being invited to the board of directors are often the same factors that lead to unrealistic optimism.

<sup>18</sup> See also chapter 20 in this volume by Kent Greenfield. Jolls and Sunstein (2006) suggest, however, that a “substantive debiasing” strategy, which we describe below, might be more effective than insulation.

<sup>19</sup> In the United States, the Sarbanes-Oxley Act does something approximating this by allocating the responsibility of all auditing functions on outside directors. Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, sec. 301.

effect of both parties' choices on outcomes. One obvious legal mechanism accomplishing this would be regulation, which is effectively an insulating strategy on *all* decision-makers.

Motor vehicle regulations provide an excellent example, where researchers have found specific evidence of overoptimism bias (Svenson 1981). Since automobile torts are characterized by role-reversibility (i.e., the activity subjects one to the simultaneous risk of being a tortfeasor and a victim), it is particularly difficult to insulate against optimistic drivers; because the set of prospective tortfeasors is identical to the set of prospective victims, whatever biases affect one party will also affect the other. By regulating certain decisions (e.g., speed, vehicle maintenance, cell phone use), the law restricts the domain of choices available to all parties.

Luppi and Parisi (2013a) point out an alternative insulating strategy when both parties suffer from similarly destructive biases in a tort law context. When parties suffer from "blind spot bias," a species of overoptimism, they may be capable of adapting when given statistics about biased behavior. However, they may tend to have a "blind spot" about *their own* susceptibility to bias.<sup>20</sup> Thus, while they may continue to form biased judgments or exhibit biased behavior, they may be able to anticipate and recognize biases in *other people*. Luppi and Parisi (2013a) point out that a party suffering from blind spot bias may increase his level of precautionary care to compensate for the suboptimal precautions he anticipates from the other party. Likewise, the other party, also suffering from blind spot bias, may do the same. Thus, reducing the due care threshold in a negligence regime to account for "ordinary" levels of overoptimism may ironically incentivize both parties to exercise greater care.

Other examples of insulating strategies include bans of any sort, which straightforwardly remove alternatives from the domain of available choices by outright prohibition. In a jury context, excluding evidence when the potential to confuse or mislead outweigh the probative value of presenting it may be regarded as another clear example of an insulating strategy.<sup>21</sup>

## 2.2 Debiasing

Ideally, when parties suffering from biased judgment are given access to the accurate information, they attempt to self-correct. For example, if a student routinely budgets insufficient time to complete assignments due to overoptimism, then in the event a benevolent observer draws his attention to this pattern of behavior, he might in the future allot more time than he thinks is needed, to compensate for his known tendency to underestimate (e.g., Stiglitz 1986). By educating him about his own suboptimal

<sup>20</sup> Pronin, Lin, and Ross (2002) identified the blind spot bias, finding that subjects were resistant to applying knowledge about biases to their own behavior, regarding themselves as "less biased than average."

<sup>21</sup> FED. RULES OF EVID. Rule 401. (2011). Chapter 27 in this volume by Frederick E. Vars, for a comprehensive investigation of debiasing and insulating strategies as they relate to evidence law.

behavior (in effect reducing the deliberation cost), he no longer exhibits (or exhibits less) biased behavior; he has been debiased.

Debiasing is the process whereby biased individuals are made to be unbiased individuals. Fundamentally, all debiasing operates through the reduction of deliberation costs. Individuals often rely on heuristics to form judgments when the cost of rational deliberation is unduly high. That is, when  $P_i(x^*)$  is the optimal payoff, and  $D_R$  is the cost of rational deliberation, it may be the case that for some alternative deliberation process,  $D_H$ , the savings in deliberation cost offsets the deviation from the optimum,  $P_i(x^*) - D_R < P_i(x^o) - D_H$ .

Of course, individuals cannot know prior to deliberating how costly a deliberation process will be in a particular circumstance. They therefore rely on certain triggering facts to dictate what heuristics to employ in what circumstances. When the triggering facts instantiating heuristics are overly broad, debiasing may operate by reducing the cost of the second-order deliberation (i.e., the decision of what first-order deliberation process to employ), possibly encouraging rational deliberation (or at least rational behavior) in some subset of cases. Debiasing makes the set of second-order “triggering facts” more fine-grained, improving the application of decision-making procedures. Alternatively, debiasing may be accomplished by reducing “search costs”—a species of deliberation cost—by simply providing the information required for rational deliberation.

A point of clarification is required on this subtle point. The reduction of deliberation costs is a form of debiasing, though not directly (in a direct sense, it more closely resembles incentive alignment). However, it can nevertheless have a debiasing *effect*. By reducing deliberation costs in some subset of cases of a given type, the net benefit of applying a heuristic to such situations *in general* is reduced. Thus, in such situations, individuals will find it second-order optimal to use rational deliberation rather than a heuristic, even in those cases where deliberation costs have *not* been reduced.<sup>22</sup>

One reasonable criticism of debiasing is that its effects (to say nothing of the problems it is meant to correct) are difficult to quantify—one clear advantage of insulating strategies is that they straightforwardly and obviously reduce the effect of biases to the extent that they remove such biases from affecting consequences. However, when populations are not homogeneously biased, insulating strategies carry the significant cost of constraining the choices of unbiased individuals as well as biased individuals. Where this cost is substantial, debiasing may represent a better alternative strategy (though the two are not necessarily incompatible, and a two-pronged attack may sometimes be warranted), since it does not foreclose the choices available to individuals, but rather eliminates the presence of bias itself, targeting it in those individuals affected.

Strategies for debiasing through law may be divided into two broad categories: adjudicative and substantive. Adjudicative debiasing involves the elimination or mitigation

<sup>22</sup> More informally, even a *partial* reduction (only applicable to a handful of instances) in deliberation costs for a type of situation will encourage individuals to “break the habit” of relying on heuristics in the future, even when deliberation costs may be higher.

of biases in adjudication, whereas substantive debiasing involves the elimination or mitigation of biases in substantive areas of law.

Let us first consider adjudicative debiasing. The greatest portion of research in adjudicative debiasing focuses on juries.<sup>23</sup> As fact-finders, jurors are especially vulnerable to the biases that infect everyday reasoning about the world. For example, Hillman (2000) points out that juries may be susceptible to hindsight bias and framing effects when assessing whether a liquidated damages clause is “punitive.”<sup>24</sup>

Recall that hindsight bias leads individuals to overestimate the probability of an event occurring when evaluating it *ex post*. That is, if an event occurs, then individuals looking back at it from a later point in time tend to regard its occurrence as being more likely than it really was (or at least more likely than it would have seemed to a reasonable person *ex ante*). Thus, a liquidated damages provision, which may have seemed “reasonable” to the contracting parties *ex ante* may be deemed “punitive,” when a jury (or judge) looks back at it *ex post*.

Rachlinski (2000) suggests that hindsight bias may be mitigated by defining categories of *per se* validity or invalidity with respect to liquidated damages, transforming the adjudicative choice from a question of whether a liquidated damages clause is punitive (a decision vulnerable to hindsight bias) to a question of whether the case fits in certain predetermined categories. Interestingly, Rachlinski (2000) also suggests that while juries may be influenced by hindsight bias in assessing liquidated damages, this may actually be a form of debiasing. That is, because contracting parties often suffer from overoptimism *ex ante*, hindsight bias in courts may have the effect of “canceling out” the biases of the litigants.<sup>25</sup>

Hillman (2000) also mentions framing effects. Recall that framing effects describe phenomena where individuals respond inconsistently when a decision is presented in two different ways. For example, Tversky and Kahneman (1981) gave test subjects the cover story, “Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people,” asking two groups of test subjects to choose between two programs to combat the disease. In the first group, subjects are asked whether they would prefer (1) a disease prevention program that will save 200

<sup>23</sup> Devine et al. (2000) provide an extensive survey of empirical research on jury decision-making. See also chapter 26 in this volume by Doron Teichman and Eyal Zamir.

<sup>24</sup> Hillman’s (2000) investigation looks to behavioral law and economics to explain why courts paradoxically favor deferring to parties’ manifest intent in contract disputes, while liberally invalidating liquidated damages clauses. See Stallard and Worthington (1998) on hindsight bias in closing arguments.

<sup>25</sup> Rachlinski (2000) is not entirely clear on this point. If a liquidated damages clause objectively ought to have specified  $\bar{X}$  value of damages, and optimistic parties agree to  $\bar{X} + \epsilon$  damages *ex ante*, and juries affected by hindsight bias tend to assess damages as  $\bar{X} - \delta$ , then it is unclear why the jury’s bias should be regarded as hedging against the contracting parties’ bias. One could alternatively read Rachlinski (2000) as saying that contracting parties anticipate the possibility that a jury will suffer from hindsight bias, setting a lower (and thereby valid) liquidated damages amount than they otherwise would have.

lives; or (2) a program that has a one-third chance of saving 600 lives, and a two-thirds chance of saving no lives. In expected utility terms, the two options are equivalent; however a majority (72%) selected the first option. Contrastingly, when the second group was asked whether they would prefer (3) a program that results in 400 deaths; or (4) a program that has a one-third chance of resulting in no deaths, and a two-thirds chance of resulting in 600 deaths, a majority (78%) of respondents preferred option (4). Of course, option (1) is equivalent to option (3), and option (2) is equivalent to option (4). Thus, depending on how the question was posed, the majority “preference” seemed to change. Similarly, in deciding whether a liquidated damages provision should be viewed as excessive, juries may well be highly susceptible to how the contract language is framed.<sup>26</sup>

Rachlinski (2000) suggests that framing biases may be eliminated by abandoning the doctrinal distinction between discounts and penalties, which tends to exacerbate (or may simply be the result of) framing effects. Of course, liquidated damages represent a specialized issue in contract law. However, the liquidated damages concerns may be generalized to include many disparate areas of law. For instance, in a tort case, hindsight bias may lead juries to overestimate the probability that an accident would occur, potentially altering jurors’ assessments of whether a tortfeasor took “reasonable precautions.”<sup>27</sup> Furthermore, framing effects may play a significant role in determining nonpecuniary damage awards. McCaffery, Kahneman, and Spitzer (1995) found that survey respondents, asked to put themselves in the position of a juror, exhibited wildly divergent responses to how much compensation should be given for pain and suffering, depending on how the question was framed. When subjects were asked to determine damages on the basis of what the victim “would have demanded to willingly accept the injury,” they answered with an average figure nearly double the sum that respondents gave when asked to determine damages on the basis of what it would take to “make the victim whole.” Efforts to curb framing effects may involve careful crafting of jury instructions, which avoid framing nonpecuniary damages in terms of losses or gains (see also Sunstein, Kahneman, and Schkade 1998).

Debiasing may also be used to combat cascade effects in jury deliberation. “Informational cascades” occur when individuals under uncertainty are exposed to the opinions of others. When this happens, uninformed people often allow their judgment to be contingent upon the judgments of others. The result is that group judgment-formation often ends up converging around a position with a high degree of confidence (Sunstein 2005; Banerjee 1992). However, because juries necessarily deliberate by sequentially disclosing their opinions, the order in which jurors express their views during juror deliberation may affect the point of convergence (Sunstein 2002). These factors may work against the predictions of Condorcet’s (1785) voting theorem,

<sup>26</sup> Hillman (2000) points out, for example, how courts fallaciously distinguish between a “penalty” for late performance and a “discount” for early performance.

<sup>27</sup> Chapter 16 in this volume by Yoed Halbersberg and Ehud Guttel, discusses the application of behavioral economics to tort law generally.



which assumes that voters' opinions are independent.<sup>28</sup> Debiasing may take the form of instructions on deliberation, such as requiring each juror to articulate his or her opinion prior to discussion, to resist the effect of cascades.<sup>29</sup>

Improving jury instructions may also help to mitigate the serial position effects (a subcategory of framing effects). Murdock (1962) found that when presented with a list of information serially, individuals tend to recall the initial items on the list ("primacy effect") and final items on the list ("recency effect") better than items of information in the middle. The two effects may thereby distort how juries weigh evidence presented during trial (Kassin and Wrightsman 1979). In addition to improved jury instructions, when a judgment is contingent upon several independent issues, a case may be divided into independent proceedings on each issue,<sup>30</sup> to mitigate serial position effects.

Though juries (and judges) have been the primary targets of procedure debiasing, litigants may also suffer from biases, which lead to inefficient outcomes. For example, overoptimism may lead litigants to overestimate the strength of their cases, decreasing (and indeed possibly extinguishing) the margin for negotiating a settlement (Babcock et al. 1995). Babcock, Loewenstein, and Issacharoff (1997) find, however, that simply pointing out to a party the weaknesses in its case (thereby reducing the deliberation cost) effectively recalibrates their expectations, debiasing them. Altering court procedures to force litigants to reflect on the merits of their opponents' claims (and the weaknesses in their own) may therefore elevate the rate of settlements, reducing overpacked dockets and court costs.<sup>31</sup>

Whereas adjudicative debiasing is aimed at correcting bias in the adjudicative institutions of the law, substantive debiasing is aimed at eliminating or reducing biases "in the world," which interfere with the objectives of substantive law.<sup>32</sup> For example, in a corporate law context, firms could be required to maintain a certain ratio of outside directors to reduce the effect of overoptimism, which we discussed earlier as potentially affecting inside directors (Jolls and Sunstein 2006; Langevoort 2001).

In deciding between a property rule and a liability rule, Jolls and Sunstein (2006) suggest that courts would be wise to consider the impact of the endowment effect. Recall that the endowment effect relates to individuals' tendency to value items they own more than those they *could* own. In the presence of endowment effects,

<sup>28</sup> Luppi and Parisi (2013b) draw together these factors to investigate the effect of the interaction between cascade effects and group polarization on hung jury rates.

<sup>29</sup> Altering jury size and voting rules may also mitigate the effects of polarization and cascades—though these ought to be regarded as insulating rather than debiasing strategies.

<sup>30</sup> As prescribed by FEDERAL RULES OF CIVIL PROCEDURE, Rule 42(B).

<sup>31</sup> Robbenholt, chapter 24 of this volume, offers a more comprehensive discussion of the behavioral factors pertaining to settlement.

<sup>32</sup> For a survey of behavioral biases and substantive debiasing techniques in the field of accident law, see Luppi and Parisi (forthcoming).

individuals may suboptimally overvalue their own property, reducing the efficiency of voluntary exchanges. In such cases, Jolls and Sunstein suggest that protecting interests with a liability rule rather than a property rule may eliminate biases attributable to the endowment effect.

Jolls and Sunstein (2006) also point to agency law as a potential mechanism for debiasing. By assigning buying or selling power to agents, a firm may eliminate the bias attributable to endowment effect. Because an agent facilitates exchanges that do not involve his own property, he will be immune from (or at least comparatively less affected by) the endowment effect that influences the firm's owners.

It is useful to note that these examples of debiasing may also be regarded as insulating strategies. The Sarbanes-Oxley Act decreases the bias of a board of directors in two ways: it decreases the number of biased directors (by requiring a minimum threshold of outside directors), and it isolates certain types of judgments (those pertaining to auditing) from biased directors. Whether the board is regarded as a single collective group or a set of individuals changes whether we categorize it as insulation or debiasing. To the extent that the board, qua collective unit, is made less biased, laws requiring that a portion of the board consist of outside directors may be regarded as a debiasing strategy. However, to the extent that the board, qua aggregation of individual directors, requires certain directors to be excluded from decision-making, it may also be modeled as insulation.

Likewise, changing a property rule into a liability rule may be regarded as either debiasing or insulation, depending on how the question is framed. If the domain of choices available to the parties does not differ as between a liability and property rule, then choosing a liability rule to combat the endowment effect may be regarded as debiasing. However, if the parties' domains of choices change as between a property and liability rule, then to the extent that those altered options are the cause of the debiasing, it should instead be categorized as insulation.

And also with respect to agency remedies, it will depend on whether the agent is conceptualized as an extension of the principal or an independent (though contractually bound) actor whether we regard delegation of decision-making as debiasing or insulation. If we regard the agent as distinct from the principal, then agency should be regarded as an insulating strategy, since it removes the principal from the decision-making (i.e., it extinguishes the domain of available choices of the principal, making the principal's behavior wholly contingent upon the agent). If, however, we regard the agent as a mere extension of the principal, then agency should be regarded as a debiasing strategy, since it makes the principal's judgments less biased (because the principal's judgments on this view simply *are* the agent's judgments).

The point here is that there is some analytical flexibility as to whether an effort to debias should be regarded as a debiasing strategy or an insulating strategy. In general, we think that this potential ambiguity is unlikely to be problematic, so long as one is consistent in framing the problem.

### 3 BIASING AND LEVERAGING

---

The foregoing section assumed at the outset that privately optimal decision-making would entail optimal social welfare. That is, debiasing and insulating strategies are meant to restore (or at least get closer to) efficiency by reducing the effects of biased judgment and biased behavior. The reasoning is that eliminating or mitigating the effects of biases brings individuals closer to privately optimal choices, which brings social welfare closer to the optimum. However, let us now consider the converse circumstance, beginning with the premise that parties behaving optimally *fail* to effect a social optimum. Equivalently, let us assume the presence of market failure. Formally, market failure may be characterized by the condition where individuals' optimal choices fail to maximize social welfare:

$$S^{**} = P_1(x^{**}) + P_2(y^{**}) \neq P_1(x^*) + P_2(y^*).$$

When privately optimal behavior fails to maximize social welfare, the law may be used to influence parties who would otherwise behave rationally to make socially optimal (but privately suboptimal) choices by inducing biased judgment. Rather than regarding biases as *problems* to be rectified, they may offer potential *solutions* to entirely unrelated troubles.

Fundamentally, the idea is that when  $P_i(x^{**}) - P_i(x^0) > P_i(x^{**}) - P_i(x^*)$ , biased behavior will be socially preferable to rational behavior. Letting  $\gamma$  denote the cost of introducing bias, if  $\gamma < P_i(x^0) - P_i(x^*)$ , then *biasing* may be a useful tool for correcting other departures from social optimality.

Biasing solutions use biases to effect efficiency rather than eliminating biases to restore efficiency. Biasing may thus be thought of as the mirror image of debiasing. Biasing will also consist of two subcategories. The counterpart to debiasing strategies we call “benevolent biasing,” and the counterpart to insulating strategies we call “cognitive leveraging” (Pi 2013).

#### 3.1 Benevolent Biasing

Pi (2013) introduces the concept of biasing in the context of criminal law. Criminal activity is an externality imposed on victims by criminals. For some individuals, crime may be privately optimal; the privately optimal behavior of criminals, however, leads to socially suboptimal outcomes. Becker's (1968) seminal model of criminal deterrence suggests that prospective criminals may be deterred from committing crimes either by increasing the severity of punishment (i.e., longer sentences) or by increasing the probability of detection (e.g., hiring more police, giving prosecutors procedural advantages over defendants). By increasing the expected cost of punishment, Becker argues that the cost of criminal activity will increase and, like other market goods, its “demand” will decrease.

However, incarceration and policing are costly. Pi suggests an alternative mechanism for deterring prospective criminals—inducing a “not to commit crime” heuristic. The idea here is that even though there will always exist opportunities to commit crime, and even though the payoff of criminal activity may make it first-order optimal for some subset of the population to commit crimes, it may be *second-order* suboptimal when deliberation costs are included in the calculation.

For example, consider a prospective car thief looking to steal a car. Suppose the expected benefit of stealing the car is 10, and the expected cost of punishment is 8. Now let us say that the deliberation cost—the cost of determining whether to steal *this* car, as well as *how* to steal it—is 1. Thus, the net expected payoff of attempting to steal the car is 1, and classical (first-order) economic theory predicts the car thief will make the attempt (assuming his opportunity cost is zero).

Underlying the first-order decision (whether to steal the car) is the second-order decision *to perform a rational deliberation* when deciding whether to steal a car. The second-order deliberation cost must therefore include the first-order deliberation costs for all the cars the thief chooses to steal *and* all the cars the thief chooses *not* to steal. When a prospective thief chooses not to attempt a theft, he must still incur the cost of rationally choosing not to steal the car. Suppose it is only (first-order) rational for him to make the attempt 10% of the time. Thus, the second-order deliberation cost of rationality is 10, and the second-order benefit is 1, and it would *not* be rational to even *consider* stealing the car.

Normatively then, the objective of criminal deterrence should not be to disincentivize prospective criminals from committing particular crimes, but rather to make it second-order rational not to contemplate committing crimes in the first place—a reconceptualization of the objectives of criminal law. Importantly, this adds a third lever to the toolkit of the policymaker (and legal scholar). Becker (1968) framed deterrence as a function of the probability and severity of punishment. Pi’s (2013) model suggests that increasing deliberation costs may offer an additional mechanism for effective deterrence, which may incidentally be considerably less costly to the state.

Propagating misinformation may induce biases and alter the subjective assessment of the expected cost of punishment (though this too may be characterized in terms of increasing deliberation costs). However, this brute force method of biasing seems fraught with considerable risks (e.g., if the misinformation campaign is exposed, or if it inspires systematic skepticism, then it is plausible that the tactic will backfire).<sup>33</sup> Instead, the same factors that create problematic biases may be harnessed to induce socially beneficial biases, without the need to employ propagandistic deception. For example, policing patterns could be designed to increase uncertainty about the probability of detection (e.g., aggressively pursuing different types of criminal activity on different days of the week, instructing police to vary their patrols from day to day). Returning to the car theft example, police could invest greater resources in “bait cars” (decoy cars

<sup>33</sup> Creating disparate pools of information as between public officials and the public has been called “acoustic separation.” For a seminal account, see Dan-Cohen (1984).

set up by police to catch car thieves); if the practice were exercised with sufficient frequency, then in addition to its effectiveness in apprehending incompetent car thieves (which may yield diminishing returns as criminals learn about the practice), it may also increase the cost of rational deliberation for savvy car thieves (which will impose a fixed cost on the deliberation of whether to steal a car for every contemplated attempt).

There is a subtle point here, which is easy to lose, and which bears distinguishing. The “biasing” effect being described is *not* the mere manipulation of deliberation costs. Certainly, increasing deliberation costs will directly elevate deterrence (by increasing the total cost of committing a crime). This is true in Becker’s (1968) model (suitably modified to incorporate deliberation costs), and it is not debiasing. Rather, the debiasing is the second-order effect, which occurs when the *net* expected payoff of committing a crime is reduced to the point where it is not even worth undertaking a rational deliberation whether to commit crime (or a certain class of crime). Under a modified Becker model, a would-be criminal will rationally deliberate whether to commit a particular criminal act, weighing the expected benefits against the expected costs; and if the expected costs outweigh the expected benefits, he will have been “deterred” from committing the crime. However, under a second-order model of rationality, a “deterred” individual would not even undertake the rational analysis, because the expected cost of *contemplating* the commission of the crime (not the expected cost of *committing* it) would outweigh the expected benefit. Rather than undertaking a rational weighing of expected payoffs, he weighs the payoff of *weighing the payoffs* in such cases. This implies that manipulating deliberation costs will have a multiplier effect, resulting in deterrence far greater than predicted under a first-order modified Becker model.

In environmental law, benevolent biasing could be employed to combat overconsumption of public goods. For example, let us suppose (as seems likely to be the case) that consumers do not fully internalize the cost of energy consumption. And let us assume that consumers are unboundedly rational actors with respect to their electricity consumption. Say that an individual is leaving home to go to an important meeting, and rushes out without turning off unused appliances in his home. Let us say the private cost of running the appliances is 9 and the expected cost of arriving late to the meeting is 10. The rational choice would be to leave the lights on to avoid being late to the meeting. Let us also posit that the social cost of leaving the appliances running is 11, given the pollution externalities that are not fully captured in the price of electricity. The individual creates a net social loss.

Of course, the obvious solution would simply be to increase the price of electricity, making the consumer internalize the full cost of power consumption. However, as is well documented in the literature, the cheaper pricing policy might be preferred, given the need to provide affordable power to residents. Thus, a less costly solution may be to induce biased behavior in the consumer.

One creative solution may be to switch from postconsumption billing to a prepayment scheme, where consumers are billed for a socially optimal rate of consumption. If the consumer ends up using less power than the social optimum, then he is awarded a discount; if the consumer ends up using more, then he is charged a penalty. Note

that the award and penalty are equivalent to the normal rate of billing. So whereas consumer  $i$  would pay  $P_i^{post} = rx_i$  in a postpayment billing scheme, where  $r$  is the rate and  $x_i$  is the quantity of consumption, under a prepayment scheme, he would pay  $P_i^{pre} = r(x_i^* + (x_i - x_i^*))$ . Of course, arithmetically,  $P_i^{post} = P_i^{pre}$  and the two payment schemes are consequentially equivalent.

However, due to framing effects, the consumer may perceive penalties due to consumption above the prepayment threshold as a loss, inducing a bias toward less consumption. The important point is that the incentives have not changed. Instead, by providing a reference point, the government (or the power company in our example) will have induced biased reasoning about consumption, causing individuals to behave privately suboptimally, but socially optimally, thereby mitigating the effect of the externality.

Clearly, this sketch leaves out much detail. A more thorough evaluation of whether such a biasing tactic would be effective should account for changes in transaction costs—for example, the increased cost of billing twice (the prepayment and the penalty/award), the discounting cost of prepayment—as well as anticipating the encouragement of other cognitive illusions (e.g., the framing may increase consumption for people who would otherwise consume less than the socially optimal average). Tailoring the prepayment scheme and including more factors in the analysis of social costs may mitigate these problems. At any rate, we provide the example not as a serious proposal, but rather as a stylized example of how biasing may be used to correct market failure. We think it is a plausible tactic; however, even if it turned out to be unrealizable in the particular case, we think the strategy is generalizable and must surely be valid in other areas of the law.

Clearly, this approach will have dangers. For instance, inducing biases in otherwise rational individuals runs the risk of distorting their otherwise optimal choices. So, in our example, the framing effect may result in underconsumption (e.g., loss-averse individuals not only take extra care to turn off appliances when leaving the home, they may refrain from using them when it is useful to do so). If the consequences of underconsumption deviate further from the social optimum than the consequences of overconsumption, the policy will obviously have failed. Nevertheless, in some cases at least, benevolent biasing may equip policymakers with an additional tool in correcting behavior in the presence of other problems of suboptimal behavior.

In a nutshell, when privately rational behavior fails to maximize social welfare, inducing biases may cause individuals to act against their own private interests and contribute to an increase in aggregate social welfare toward the social optimum.

### 3.2 Cognitive Leveraging

Let us now consider the biasing counterpart to insulation. Recall that with insulating strategies, the fundamental goal is to *decrease* the domain of available alternatives for biased decision-makers, with the aim of reducing their ability to affect outcomes. With cognitive leveraging, the objective is to *increase* the domain of available alternatives



for apparently unbiased decision-makers, with the aim of exposing a bias previously irrelevant to outcomes.

For example, sticking to the analytically convenient example of power consumption and pollution externalities, consider the following policy: individuals may prepay for as much electricity as they expect to use at the regular rate. However, any consumption beyond the prepayment amount will result in overage penalties (the “penalty” here may be a genuine penalty—that is, it may be greater than the regular rate). Notice that this policy would not affect an unboundedly rational individual’s payoffs. If individual  $i$  paid  $rx_i$  under the ordinary postpayment scheme, then  $i$  would pay the same under the prepayment scheme, provided he accurately predicted how much electricity he would use during the month. But if individual  $i$  suffers from overoptimism, then he may underestimate his future energy consumption, precommitting himself to a lower (possibly socially optimal) level of usage than what may have been privately optimal.

Careful attention needs to be given to the distinction between leveraging, biasing, and incentive alignment. Incentive alignment effects socially optimal choices by altering the private *payoffs* of individuals. Notice that incentives were not altered in either of our examples of benevolent biasing or cognitive leveraging. An unboundedly rational individual would not alter her behavior under either of the prepayment alternatives, and her payoffs would remain absolutely identical. With benevolent biasing, rational actors are presented with information in such a way that the cost of rational deliberation increases; they therefore select first-order suboptimal choices. That is, rational actors are transformed into biased actors. With cognitive leveraging, individuals’ biases are already present in the judgments they form—biases, which would bring their behavior closer to the social optimum. However, they are unable to act on those biased judgments, because their payoffs are naturally “insulated” from their biased judgments. In our example, *ex ante* overoptimism about one’s expected electricity usage (or wastefulness) is insulated in a postpayment system. Switching to a prepayment scheme, whereby consumers are essentially forced to bet on how much electricity they will use in a month, uncovers the preexisting bias. Unlike biasing, it does not create a bias where none existed, but rather exposes consequences to already existing biased judgments—it “deinsulates” outcomes.

Of course, the inducement of biases to effect efficiency is not new. Jolls and Sunstein (2006) provide several examples, where biases may be harnessed to combat inefficiency (see also chapter 28 by Sunstein in this volume). When straightforward debiasing—reducing deliberation costs—fails to counteract overoptimism bias (Weinstein and Klein 2002), Jolls and Sunstein suggest the inducement of countervailing biases to effect a sort of canceling out. Indeed, most of the examples they give (and which we discussed in the previous section) may be better regarded as “double biasing” rather than legitimate debiasing.

For example, cigarette smokers may suffer from unrealistic optimism about the increased probability of lung cancer due to smoking. Policymakers may counteract smokers’ overoptimism bias by using availability bias to increase smokers’ sensitivity to the risk of smoking. Recall that availability bias describes a subjective overvaluation



of specific information (relative to general or statistical information). Thus, when smokers are exposed to information about particular cases of other smokers developing lung cancer, they are more likely to quit or reduce their own consumption of cigarettes (Sloan, Taylor, and Smith 2003). The idea is that the unrealistic overvaluation of particular instances offsets the unrealistic undervaluation of the probability that one will develop cancer oneself.

However, Jolls and Sunstein (2006) only consider the inducement and leveraging of biases as a corrective against *other biases*—a special case of the biasing approach we have described here. There is reason to suspect that Jolls and Sunstein’s use of biasing and leveraging—that is, using biases to fight biases—may be the least promising application of these principles. First, it seems intuitively likely that quantifying the effects of biases may pose considerable practical difficulties, and if so, then our recognition of suboptimal biases may often be approximate at best. Deploying corrective biases to counteract suboptimal biases compounds the problem of imprecision, and the dangers of overshooting are likely to compound as individuals depart further from rational deliberation. Just as we should be especially careful in evaluating whether two wrongs can make a right morally, we should likewise be especially cautious about concluding that two biases can make an individual unbiased behaviorally.

Moreover, the interaction of multiple biases, pushing individuals in different directions, may not even be additive. That is, if bias *A* tends to increase consumption, and bias *B* tends to decrease consumption, there is no reason to assume that inducing *B* to correct for *A* will result in a canceling out of the two distortions. Jolls and Sunstein implicitly assume that if  $-1 \times A \approx B$ , then  $x^* + A + B \approx x^*$ , which is sound arithmetic, but may not be a good model of how biases interact. For instance, representativeness bias does not seem to interact with rational deliberation in a merely additive fashion. Rather, it seems to cancel out rational deliberation entirely (Tversky and Kahneman 1974). Inducing a new bias may not only offset, but may entirely corrode, the effects of the old one. That is, when bias *B* is introduced to counteract bias *A*, it may be that individual behavior will simply be driven by *B* rather than *A*, instead of being driven by the *sum* of *B* and *A*.

Thus, it may be that biasing and leveraging strategies may be better deployed to correct other forms of suboptimal behavior in circumstances where there exists *no* prior biased behavior (though in the case of leveraging, there will be prior biased judgment that fails to cash out in behavior). This may be beneficial in cases where privately optimal behavior fails to be socially efficient.

We conclude this section with a taxonomical note. In our exposition of debiasing, we followed Jolls and Sunstein (2006) in distinguishing between adjudicative and substantive debiasing.<sup>34</sup> It may be tempting therefore to attempt to apply those categories to biasing and leveraging. We think, however, that there is little room for biasing or leveraging in an adjudicative context. Even if biases are additive, we should be extremely

<sup>34</sup> Jolls and Sunstein (2006) use the term “procedural” rather than “adjudicative.”

skeptical of policies deliberately designed to make courts *more biased*, even if it is to counteract the effects of other preexisting biases.<sup>35</sup>

## 4 CONCLUSION

---

In this chapter, we provided a broad overview of debiasing and insulating strategies, briefly discussing both their theory and applications. We also introduced the two logical counterparts to debiasing and insulating: benevolent biasing and cognitive leveraging, expanding on the work of Pi (2013). We have provided examples from various areas of law both to convey an intuitive understanding of how the theory has (or might be) employed, and also to illustrate the breadth of the applicability of debiasing and biasing in law.

The relationship between the four forms of social engineering may be summed up (with some admitted oversimplification) thusly: if biases result in behavior deviating from the social optimum, then the law may be used to eliminate or mitigate the bias (debiasing). However, if biased individuals are resistant to debiasing, or if debiasing is comparatively costly, then the law may be designed around biased individuals, to eliminate or mitigate the effects of bias on outcomes (insulation). On the other hand, if rational judgment results in behavior deviating from the social optimum, then the law may be used to induce biases, which nudge behavior close to the social optimum (benevolent biasing). When individuals are already biased, then the law may be designed to depend on those biases to correct some other form of inefficiency, which may arise when individuals are rational (cognitive leveraging).

These topics promise to be robust areas for new research for many years, giving behavioral economics a practical significance, about which legal scholars had once expressed skepticism (e.g., Hillman 2000). It is interesting to observe that the study of behavioral economics began with observations of real-world deviations from the predictions of rational choice theory (Simon 1955; Allais 1953). The innovation (and controversy) of behavioral economics was to raise doubts on rational choice theory as an accurate description of human behavior and to question its predictive value and usefulness in policy analysis—a positive, not normative, controversy. Thus, to the extent that welfare maximization is a valid goal, behavioral economics makes no contrary normative claim. The absence of a prescriptive element therefore left many scholars questioning the practical significance of behavioral economics (Hillman 2000; Rachlinski 2000). Debiasing and biasing go far in answering this concern, giving a normative bite to the positive analysis of psychologists and behavioral economists.

<sup>35</sup> However, it is conceivable that in some cases, where counteracting biases are of the same type, such approaches may be promising. For example, biases due to framing effects may be mitigated or eliminated by instituting procedural rules that alter a judge's or juror's reference point.

## REFERENCES

- Allais, M. 1953. Le comportement de l'homme rationnel devant le risqué: Critique des postulats et axiomes de l'école américaine. *Econometrica* 21:503–46.
- Alicke, M. D., and O. Govorun. 2005. The better-than-average effect. In *The self in social judgment*, M. D. Alicke, D. A. Dunning, and J. I. Krueger. New York: Psychology Press, 85–106.
- Babcock, L., G. Loewenstein, and S. Issacharoff. 1997. Creating convergence: Debiasing biased litigants. *Law and Social Inquiry* 22:913–25.
- Babcock, L., G. Loewenstein, S. Issacharoff, and C. Camerer. 1995. Biased judgments of fairness and bargaining. *American Economic Review* 85:1337–43.
- Banerjee, A. V. 1992. A simple model of herd behavior. *Quarterly Journal of Economics* 107(3):797–817.
- Becker, G. S. 1968. Crime and punishment: An economic approach. *Journal of Political Economy*. 76/2:169–217.
- Cantril, H. 1938. The prediction of social events. *Journal of Abnormal and Social Psychology* 33:364–89.
- Marquis de Condorcet. [1785] 1976. *Condorcet: Selected writings*. Indianapolis: Bobbs-Merrill.
- Dan Cohen, M. 1984. Decision rules and conduct rules: On acoustic separation in criminal law. *Harvard Law Review* 97/3:625–77.
- Devine, D. J., L. D. Clayton, B. B. Dunford, R. Seying, and J. Pryce. 2000. Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy, and Law* 7/3:622–727.
- Elster, J. 1985. *Sour grapes: studies in the subversion of rationality*. New York: Cambridge University Press.
- Fischhoff, B. 2003. Hindsight  $\neq$  foresight: The effect of knowledge on judgment under uncertainty. *Qual Saf Health Care* 12:304–12.
- Fischhoff, B. 2007. An early history of hindsight research. *Social Cognition* 25:10–13.
- Hillman, R. A. 2000. The limits of behavioral decision theory in legal analysis: The case of liquidated damages. *Cornell Law Review* 85:717–38.
- Jolls, C., and C. R. Sunstein. 2006. Debiasing through law. *Journal of Legal Studies* 35:199–241.
- Kassin, S. M., and L. S. Wrightsman. 1979. On the requirements of proof: The timing of judicial instruction and mock juror verdicts. *Journal of Personality and Social Psychology* 37:1877–87.
- Kahneman, D., J. L. Knetsch, and R. Thaler. 1991. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives* 5(1):193–206.
- Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47/2:263–92.
- Langevoort, D. C. 2001. The human nature of corporate boards: Law, norms, and the unintended consequences of independence and accountability. *Georgetown Law Journal* 89:797–832.
- Lee, J. Y. 1991. *Garrison Keillor: A voice of America*. Jackson: University Press of Mississippi.
- Lund, F. H. 1925. The psychology of belief. *Journal of Abnormal and Social Psychology* 20:174–96.
- Luppi, B., and F. Parisi. 2013a. Optimal liability for optimistic tortfeasors, Minnesota Legal Studies Research Paper No. 13-06.
- Luppi, B., and F. Parisi. 2013b. Jury size and the hung-jury paradox, Minnesota Legal Studies Research Paper No. 12-01.

- Luppi, B., and F. Parisi. Forthcoming. Behavioral approach to tort law. In *Handbook of behavioral law and economics*, J. Teitelbaum and K. Zeiler. Edward Elgar.
- McCaffery, E. J., D. J. Kahneman, and M. L. Spitzer. 1995. Framing the jury: Cognitive perspectives on pain and suffering awards. *Virginia Law Review* 81:1341–420
- Morewedge, C. K., L. L. Shu, D. T. Gilber, and T. D. Wilson. 2009. Bad riddance or good rubbish? Ownership and not loss aversion causes the endowment effect. *Journal of Experimental Social Psychology* 45:947–51.
- Murdock, B. B., Jr. 1962. The serial position effect of free recall. *Journal of Experimental Psychology*. 64:482–88.
- Pi, D. 2013. Using bounded rationality to fight crime. Available at SSRN: <http://ssrn.com/abstract=2214504>.
- Pronin, E., D. Y. Lin, and L. Ross. 2002. The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin* 28:369–81.
- Rachlinski, J. 2000. The “new” law and psychology: A reply to critics, skeptics, and cautious supporters. *Cornell Law Review* 85:739–66.
- Shavell, S. M. 1980. Strict liability versus negligence. *Journal of Legal Studies* 9:1–25.
- Simon, H. 1955. A behavioral model of rational choice. *Quarterly Journal of Economics*. 69:99–188.
- Sloan, F. A., D. H. Taylor, and V. K. Smith. 2003. *The smoking puzzle: Information, risk perception, and choice*. Cambridge, MA: Harvard University Press.
- Stallard, M. J., and D. L. Worthington. 1998. Reducing the hindsight bias utilizing attorney closing arguments. *Law and Human Behavior* 22:671–83.
- Stiglitz, J. E. 1986. *Economics of the public sector*. New York: Norton.
- Sunstein, C. R. 2002. The law of group polarization. *Journal of Political Philosophy* 10:175–95.
- Sunstein, C. R. 2005. Group judgments: Deliberation, statistical means, and information markets. *New York University Law Review* 80:1–88.
- Sunstein, C. R., D. Kahneman, and D. Schkade. 1998. Assessing punitive damages (with notes on cognition and valuation in law). *Yale Law Journal* 107/7:2071–153.
- Svenson, O. 1981. Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica* 47:143–48.
- Thaler, R. J. 1980. Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization* 1:39–60.
- Thaler, R. J. 1985. Mental accounting and consumer choice. *Marketing Science* 4:199–214.
- Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science*, new series, 185/4157:1124–31.
- Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science*, new series, 211/4481: 453–58.
- Tversky, A., and D. Kahneman. 1991. Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics* 107:1039–61.
- Weinstein, N. D. 1980. Unrealistic optimism about future life events. *Journal of Personality and Social Psychology* 39:806–20.
- Weinstein, N. D. and W. M. Klein. 2002. Resistance of personal risk perceptions to debiasing interventions. In *Heuristics and biases: The psychology of intuitive judgment*, T. Gilovich, D. Griffin, and D. Kahneman. New York: Cambridge University Press, 313–23.
- Wengler, L., and A. Rosén. 2000. Measuring optimism-pessimism from beliefs about future events. *Personality and Individual Differences* 28:717–28.

## CHAPTER 7

---

# ALTERNATIVE BEHAVIORAL LAW AND ECONOMICS

---

GREGORY MITCHELL

BEHAVIORAL law and economics (“BLE”) emerged as a cohesive movement in the 1990s with the publication of several articles attacking orthodox law and economics (“L&E”) as behaviorally naive (see, e.g., Hanson and Kysar 1999a, 1999b; Jolls, Sunstein, and Thaler 1998; Korobkin 1998; Korobkin and Guthrie 1994, 1997; Korobkin and Ulen 2000; Langevoort 1992, 1995; Rachlinski 1996, 1998, 2000). In addition to a common foe in the form of L&E’s rational, self-interested actor, the foundational BLE articles shared a perspective on who should replace L&E’s star. BLE’s stand-in was decidedly less attractive cognitively than L&E’s perfectly rational actor, hobbled as BLE’s actor was by many biases in judgment and inconsistent choices that make him an easy mark, but his motivational makeup was more appealing, at least to everyone but economists, because he was willing to sacrifice himself for the good of the group or to be fair. That motivational plus was offset, however, by BLE’s actor lacking the iron will of L&E’s actor, who can indefinitely delay gratification in order to maximize utility. This imperfect-and-impatient-but-less-selfish actor came to BLE from psychology studies falsifying the postulates of rational choice theory, and in particular from the portrait of human judgment and decision-making advanced in seminal work by Daniel Kahneman and Amos Tversky.<sup>1</sup>

The foundational BLE articles also shared a methodological approach. Because these works were, for the most part, nonempirical, BLE adherents accepted the methodology

<sup>1</sup> Important collections and overviews of research by Kahneman and Tversky, and related research by others, include Gilovich, Griffin and Kahneman (2002), Kahneman (2011), Kahneman, Slovic, and Tversky (1982), and Kahneman and Tversky (2000). Kahneman and Tversky’s work is part of the research program begun by Ward Edwards, often called behavioral decision theory following the title of Edwards’s 1961 chapter in the *Annual Review of Psychology* (Edwards 1961). Behavioral decision theory examines whether actual judgments and decisions match predictions based on theories of optimal judgment and choice developed by economists and game theorists (see Dawes 1985; Edwards 1954, 1961). Many psychologists other than Kahneman and Tversky have made important contributions to this area of research, but the theories and metaphors developed by Kahneman and Tversky have been most influential within BLE.

of the psychology studies on which they relied, and used these studies to build simple, nonmathematical models of how the average individual supposedly deals with a variety of judgments and choices with legal implications, from decisions to engage in criminal behavior or accept a consumer product warranty to voting decisions to the processing of evidence at trials to arrive at judgments of guilt or liability. The typical BLE article asks how our view of the law, legal institutions, or markets might change if we replace L&E's perfectly rational and selfish actor with someone who predictably exhibits some sort of irrational or unselfish behavior. Would this alternative view provide new justification for existing laws, as Rachlinski's (1998) work on the hindsight bias did for certain elements of tort law, would it justify new laws or changes to existing laws, as with Sunstein, Kahneman, and Schkade's (1998) work on punitive damages, or would it disturb long-standing beliefs about a market's self-regulating nature, as with Langevoort's (1992) work on the efficient market hypothesis?

BLE thus constituted more than an attempt to make L&E responsive to empirical evidence; it advanced a particular view of human rationality and motivation and embraced a particular methodology for developing alternative models of legal judgment and choice. BLE also embraced a particular conception of the law, but not one at odds with L&E, in which the law sets prices for different behaviors or imposing costs along a continuous scale that should be counted within a consequentialist, cost-benefit framework (e.g., Korobkin 2011). How people relate to the law, and what people want from the law, are little studied by BLE scholars.

This chapter considers the psychological, methodological, and normative paths taken by BLE and alternative paths that BLE might have taken, and might still take. The counterfactual BLE imagined here stresses the B in BLE, with behavioral approaches to legal problems being the focus rather than advocacy of any particular basic model of human cognition and motivation to compete with L&E's dominant model. This change in focus would give priority to empirical studies in which particular legal institutions and specific legal tasks are simulated or studied *in situ* rather than to studies of abstract and general judgment and decision-making problems that may provide more theoretical bang but have less clear applied payoff in specific legal contexts.

## 1 RATIONALITY AND MOTIVATION

---

The cognitive and motivational assumptions of L&E are well known and need little description here: the perfectly rational actor of L&E models maximizes his utility, meaning in simple terms that he adjusts his subjective probabilities appropriately in response to new information and makes consistent and logical choices (see McFadden 1999; Posner 2010). L&E's rational choice model provides substantial benefits. Like any good social scientific model, it enables predictions about how individuals and human organizations will behave. It also provides prescriptions about how individuals and organizations should behave if they want to maximize utility. With the help



of L&E's model, legislators or judges can formulate statutes or common-law doctrine designed to encourage or deter certain behaviors by adjusting the consequences of different courses of action and by ensuring that the information needed to make utility-maximizing choices is made available to the decision-maker. Of course, if individuals regularly depart from how L&E predicts they will or should act, then law fashioned after L&E's model may be much less effective than the legislators and judges hoped.

Enter BLE. BLE founders proposed that individuals do in fact regularly violate the predictions of rational choice theory and cannot conform themselves to the postulates of rational choice theory even if they would ideally like to do so. That proposal itself was not particularly significant, for many economists and L&E scholars had long conceded that predictions from rational choice theory often failed to describe the actions of many people, but BLE founders importantly claimed that these departures were predictable because individuals act, in McFadden's (1999) terms, not like "Chicago man" (short for the perfectly rational actor touted by Gary Becker and L&E scholars) but like "K-T man" (short for the model human built from the work of Kahneman and Tversky) (see also Prentice 2003). K-T man relies on cognitive heuristics, or mental rules of thumb, rather than careful application of the laws of probability and logic, to make judgments, and he constructs preferences on the spot rather than consulting a stable set of preferences, opening himself up to fleeting influences of the situation, such as attention-grabbing information, the manner in which choices are presented, and how filling his breakfast was. It is this reliance on cognitive heuristics and on-the-spot preference construction that leads to predictable biases in judgment and choice and predictable departures from the ideals laid out in L&E.

The K-T man of BLE is not a complete man, however. BLE's portrait of human judgment and decision-making fashioned after K-T man is only a sketch of basic abilities, inclinations, and temperament. To make a prescription for a particular legal judgment or choice, the BLE scholar must put more content into the picture, and that is where BLE runs into the problem of how to incorporate the many empirical findings on judgment and decision-making that can point in opposing directions.<sup>2</sup> As the BLE scholar adds detail to the portrait of K-T man to insert the cognitive flaw or motivational characteristic seen as relevant to the legal judgment or choice of interest, other imperfections and motivations are painted over to avoid complicating the prescriptions or explanations

<sup>2</sup> The Wikipedia entry on biases in judgment and decision-making presently lists 92 "decision-making, belief and behavioural biases," 27 "social biases," and 48 "memory errors and biases" ([http://en.wikipedia.org/wiki/List\\_of\\_biases\\_in\\_judgment\\_and\\_decision\\_making](http://en.wikipedia.org/wiki/List_of_biases_in_judgment_and_decision_making)). Many of the listed biases can point judgments and decisions in opposing directions (e.g., when will the "hostile media effect" counter attentional bias associated with news stories, or when will stereotypes about a group counter attributional biases?), and some may combine to create particularly insuperable problems (e.g., multiple mechanisms may contribute to the overweighting of negative outcomes). Without situated empirical research, it is difficult to say, *ex ante*, which combinations of biases will be at work with what net effects.



to be offered. For instance, when considering how the law might deal with K-T man's irrational concern about dread risks (low probability catastrophic events), the over-optimism of K-T man that should counter the influence of at least some dread risks can be acknowledged but not allowed to silence the dread-risk alarm (e.g., Sunstein 1998) (for additional discussion and examples, see Hastie 2001; Teichman 2011). This selective approach to irrational tendencies poses little risk to scholarly reputations and movement building, for BLE scholarship rarely engages in actual empirical competitions with L&E (and vice versa) (though this may be changing; see chapter 5 by Engel in this volume on empirical behavioral studies). So long as demonstrations of judgmental bias, decision inconsistency, short-sightedness, and other-regarding behavior proliferate within psychology, the shortcomings and quirks of BLE's actor will proliferate and the different ways that legal judgments and decisions can be analyzed from an armchair-bound but behaviorally informed perspective will as well.

In short, there is one way to be rational within L&E, but there are many ways to be predictably irrational within BLE. This multiplicity of irrational possibilities fundamentally complicates the prescriptive program of BLE, and puts BLE at risk of being more predictively inept than L&E. No solution to this fundamental problem has been offered within BLE, which continues to hold to an abstract guiding model that gets modified as needed to address seemingly any and every legal topic. Despite the pronounced intention that BLE be faithful to the behavioral evidence wherever it goes (e.g., Jolls, Sunstein, and Thaler 1998), the revealed preference of BLE adherents seems to be in favor of an abstract and general model of human behavior that can compete prescriptively, in terms of scope, with L&E's model. Perhaps the calculus is that the imperfect K-T man of BLE is far from descriptively perfect but still a good bit better than the perfect Chicago man of L&E (that seems to be the view taken, for instance, in Korobkin 2011 and Prentice 2003). Coming to some sort of overall assessment of the relative accuracy and utility of the competing models seems near impossible, but if the endgame was merger with L&E rather than continued competition, that overall assessment is unnecessary: what matters is figuring out how best to model behavior in particular circumstances to formulate more effective laws (cf. Teichman 2011).

That *behavioral* project is unlikely to succeed, however, so long as BLE scholars tweak their basic model as needed to portray individuals as irrational and in need of greater help or control whatever the circumstance—so long as, in other words, empirical evidence is recruited to support an assumption of irrational behavior rather than having the empirical evidence drive predictions about how people will behave in particular circumstances. BLE scholars may object to this characterization of how the process works, and certainly some BLE scholars have pushed back against reflexive paternalism. Political commitments and values may lead some BLE scholars to a proregulation ideology of irrationality (and lead some L&E scholars to a promarket ideology of rationality), but others may adopt a default assumption of pervasive irrationality through a faithful reading of the psychological research, which over the last few decades has conveyed quite a negative view of human rationality. Whether that view is well founded is the subject of debate among psychologists (see, e.g., Krueger and Funder 2004; Krueger

2012). Whatever the intention of the founding scholars (Jolls, Sunstein, and Thaler 1998, for instance, indicated that it was their intent to encourage anti-antipaternalism in response to L&E's antipaternalistic attitude), BLE quickly became a movement to advance a particular view of human cognition and motivation at odds with the view advanced by L&E rather than a movement based on the empirical study of legal judgment and decision-making. The aspiration to make the economic analysis of law more behavioral certainly succeeded if we judge success by the number of citations to research by Kahneman and Tversky before versus after the emergence of BLE,<sup>3</sup> or if the metric is policy influence (Osborne and Thaler 2010; Sunstein 2011). If the measure is whether we better understand and can better predict how particular laws actually affect perceived consequences and behavior, then success is much less clear.

A positive signal of BLE's commitment to empirical evidence rather than to the perpetuation of a particular view of human nature is the recent change to the broad sketch of judgment and decision-making offered by those working within the BLE framework. Originally K-T man was driven almost exclusively by automatic, intuitive, heuristic-based responses to stimuli, and the possibility that reflection or deliberation could debias judgments or decisions was affirmatively dismissed (see Mitchell 2002a, 2002b, 2009). The K-T man currently found in BLE is still driven primarily by heuristic responses to stimuli (the "System 1" response), but in the new model slower, more reflective responses (the "System 2" response) can win out with greater frequency (e.g., Baker 2011; Rachlinski 2012; Richardson and Goff 2012). This shift to a greater role for System 2 processing follows the accumulation of much evidence qualifying the original work of Kahneman and Tversky and follows, significantly, Kahneman's embrace of this revised view (Kahneman 2011; Kahneman and Frederick 2002, 2005). Particularly important in this shift was work by Stanovich and West on individual differences in rationality (see, e.g., Stanovich 1999; Stanovich and West 2008).

As positive as this shift is in demonstrating BLE's engagement with developments in behavioral decision theory, it further complicates the prescriptive program of BLE. For now BLE scholars must determine who will predictably rely on System 1 responses versus System 2 responses and when they will do so (i.e., are some individuals inherently more reflective, and do some situations tend to activate more reflective responses across persons?). This shift means that some groups of individuals and situations may be more appropriately modeled after L&E's rational actor, while others will fit one of the variants of the BLE model. This shift also means that BLE scholars should consider whether the law or the market provides a better means of sorting more from less reflective individuals and encouraging System 2 responses.

And as should be clear, even with this shift BLE retains its basic commitment to the portrait of human judgment and decision-making laid out by Kahneman and

<sup>3</sup> As of June 15, 2014, the journals and law review database of Westlaw contained 310 citations to "Kahneman and Tversky" in works published before January 1, 1995, and 2,139 citations in works published on or after January 1, 1995.

Tversky. That portrait emphasizes domain general heuristics (domain general in terms of the content of the judgments—encompassing both moral and nonmoral domains, according to Sunstein [2005]—and in that they are supposedly used across the situations in which judgments occur) and a limited set of contextual influences on choice (emphasizing in particular the framing of choices as losses or gains around reference points and the proximity of choice probabilities to zero). While the K-T view has been extremely influential and productive within behavioral decision theory, it is only one among several views on how people make judgments and decisions, on how psychological and environmental variables interact, and, as discussed in the next part, on how to study judgment and decision-making. Moreover, the research program advanced by Kahneman and Tversky is primarily a basic science program and not an applied science program in the sense of developing solutions for concrete problems. Certainly this work can lead to the sort of prescriptive interventions Baron (2012) discusses, but most behavioral decision theory research does not focus on testing interventions in real problem settings. It seeks to test theories of how people process information and make judgments and decisions under uncertainty; it does not struggle with the complications of implementing its findings. If alternative approaches enjoy greater theoretical or applied success in particular domains of legal relevance, then perhaps K-T man is not the best sketch to organize BLE, or it is not the only sketch that should compete with Chicago man. Many alternative sketches could be drawn. I identify some alternative perspectives to demonstrate that the particular perspective chosen by BLE founders was not foreordained. Each of these alternative perspectives holds to the view that rationality is bounded, and is thus at odds with L&E, but each configures the bounds differently.

One alternative, which has already been discussed a bit by BLE scholars (e.g., Rachlinski 2011, 2012) but with much less frequency than the work of Kahneman and Tversky, is offered by Gigerenzer and others who argue that people rely on “fast and frugal” heuristics to make judgments and decision (e.g., Gigerenzer 2000; Gigerenzer and Selten 2002). Fast and frugal heuristics differ from the heuristics conceived by Kahneman and Tversky in two key respects: fast and frugal heuristics are (1) domain-specific adaptations to the environment (2) that lead to better results than even deliberate, rational calculations when optimization is either not possible or the optimal strategy is unknowable. As Gigerenzer (2006) puts it, fast and frugal heuristics provide solutions to problems, whereas the heuristics proposed by Kahneman and Tversky create problems (in the form of biases that lead to errors) (cf. Kahneman 2011). Kelman (2011) provides a detailed comparison of these contrasting views of heuristics. For present purposes, suffice it to say, first, that both heuristics-based research programs have enjoyed considerable empirical support and presently neither program can declare victory,<sup>4</sup> and, second, that a BLE based on the fast and frugal view of heuristics would likely look very different than the current BLE (Rachlinski 2011). The fast and

<sup>4</sup> Each research program provides important insights and can explain some results that the other cannot; likewise, each program has suffered its share of setbacks and has its dissenters. Given the breadth of each program and the maneuvers that can be made to protect the hard core of each

frugal perspective emphasizes the environmental cues available in any given task, is not terribly concerned about errors that occur in the supposedly artificial setups used by Kahneman, Tversky, and those following their lead, and is much more sanguine about creating environments (including through the reform of laws and legal institutions) that can lead to better judgments and choices (Gigerenzer 2006; see, e.g., Sedlmeier 1999).<sup>5</sup>

Another perspective that dissents from the K-T view that errors arise from biased heuristic processing of information attributes a variety of judgmental errors, from reliance on inaccurate stereotypes of others to overconfidence about one's own abilities, to biased samples (e.g., Fiedler 2000; Fiedler and Juslin 2006). In this view, the problem (or a key part of the problem) is biased information instead of biased processing of information. If individuals base their judgments or decisions on unrepresentative samples of information, due either to the way information is found in the environment or to a biased but adaptive information search (Denrell and Le Mens 2012), then we should expect certain errors to occur and persist (particularly errors that arise from avoidance of new information and experiences, such as stereotyping, illusory correlation, overconfidence, and the overweighting of negative information and the prospect of a loss). BLE scholars sometimes discount the possibility that ensuring the disclosure of information can lead to better decisions (e.g., Paredes 2003); the biased-samples perspective suggests that laws that lead to more active information search and counteract biased samples and biased sampling tendencies may be an effective way of attacking some market failures.

A more general theory of cognition with implications for legal judgment and decision-making is fuzzy-trace theory, which arose from false memory research (see Reyna 2012; Reyna, Lloyd, and Brainerd 2003). Encoding and retrieval of memories are front and center in fuzzy-trace theory, and in other recent approaches to judgment and decision-making (e.g., Hilbert 2012). The key premise of fuzzy-trace theory is that people encode and rely on both gist and verbatim information about stimuli, with the former based on essential meaning(s) drawn from stimuli and the latter based on an exact representation of stimuli. These different types of memories often lead to different judgments and decisions, with gist memories tending naturally to exert a greater impact but verbatim memories being utilized when a task or situations calls for greater

program, it is unlikely either program, overall, will suffer an empirical knockout. A more fruitful approach is to examine success domain by domain, but even within specific domains the outcome may not be clear. For instance, an adversarial collaboration design to adjudicate between competing heuristic accounts of conjunction effects, as originally observed by Tversky and Kahneman (1982), reached an inconclusive outcome (Mellers, Hertwig, and Kahneman 2001).

<sup>5</sup> The fast and frugal heuristics program is similar in important respects to social judgment theory, which grew out of Brunswik's (1956) probabilistic functionalism and its lens model, which focuses the researcher on how individuals aggregate information from environmental cues to arrive at judgments (see Goldstein 2004). Gigerenzer's work was also influenced by Brunswik but was not as faithful an adherent to Brunswik's ideas. Social judgment theory, like the fast and frugal program, emphasizes the role of the environment in determining how well judges perform, and it emphasizes accuracy over procedural rationality in the form of consistency and coherence (see Dunwoody 2009).

detail or precision. A reliance on gist meaning may produce judgments and decisions that violate the predictions of rational choice theory based on verbatim representations of a judgment or choice scenario, or verbatim memories may interfere with tasks that should be decided via gist (see Reyna et al. 2003, for a fuller discussion of the sources of rationality errors under fuzzy-trace theory). An important point from fuzzy-trace theory is that over time people may develop better understandings of situations and of their preferences, allowing gist-based reasoning to handle complex problems very effectively and to produce satisfying outcomes because gist memories tend to be more stable over time and easier to manipulate mentally than verbatim memories (Reyna 2012). In this respect, fuzzy-trace theory resembles the fast and frugal view in that both reject the view that more, and more analytical, information processing necessarily leads to better outcomes.

Changing the key source of irrationality from heuristics and biases to gist- and verbatim-based reasoning would change possible prescriptions (e.g., under fuzzy-trace theory, alterations in the task can lead people to rely on verbatim memories, an appreciation of how gist representations are formed may lead to changes in how risks are communicated (e.g., Stone et al. 1994), and gist-based preferences are more stable than the constructed-preferences view of K-T man and BLE). Fuzzy-trace theory reserves a place for heuristic-driven, or rule-based, reasoning, where broad principles develop over time and reside in long-term memory for use on qualitative gists, but the heuristics differ from those emphasized by Kahneman and Tversky and can come into conflict with one another. These conflicts, rather than processing biases or errors, sometimes explain supposed inconsistencies in choices or judgments. For instance, a simple principle that more is better than less may come into conflict with an equity principle under which each person should have the same chance of survival in the Asian disease problem. The recognition of trade-offs between general principles in specific cases is a sign of competence, not incompetence, in fuzzy-trace theory (Reyna et al. 2003).

Furthermore, embracing fuzzy-trace theory could fundamentally alter how we conceive of the law and legal tasks. If priority is given to gist meaning and to higher-level representations of quantitative information (nominal over ordinal over interval representations), then it may be unrealistic and counterproductive to assume that fine distinctions in sanctions and rewards will be meaningful to actors (e.g., will the deterrence benefits of a five-year increase in a mandatory minimum sentence justify the added incarceration costs?), or to those making awards after harms have occurred (Hans and Reyna 2011). If there is a preference for fuzzy processing, and broad principles applied to gist meanings are more likely to drive judgments and decisions than detailed consideration of verbatim representations of specific probabilities, payoffs, and risks, then the law may waste tremendous resources developing and imposing laws that expect criminals, consumers, judges, jurors, and others to make fine-grained distinctions that are psychologically and behaviorally meaningless. If correct, this means that at times the law may be “too potent” from an efficiency perspective. That is, people may give too much weight to legal proscriptions or prescriptions and engage in excessively cautious behavior.

Another approach, the adoption of which would lead to a very different perspective on legal institutions, is found in what Klein (1998) calls the naturalistic decision-making research program. Contrary to the BLE view that all individuals fall prey to predictable irrationalities (e.g., Guthrie, Rachlinski, and Wistrich 2001; Rachlinski 2000; but see Rachlinski, Wistrich, and Guthrie 2006; Wistrich, Guthrie, and Rachlinski 2005), naturalistic decision-making researchers have demonstrated that, in a number of domains with identifiable conditions, expert judgments reliably outperform the judgments of nonexperts (see Pliske and Klein 2003—outside those domains, expert judgment can be spectacularly bad and misleading [Tetlock 2005]). We should have greater confidence in expert judgments when experts operate in domains where valid cues about the causal and statistical structure of the environment can be discerned and learned (i.e., the expert receives clear and quick feedback on the accuracy or efficacy of her judgments so that she can learn which cues and patterns signal important regularities in situations) (see Kahneman and Klein 2009). This perspective simultaneously encourages greater reliance on experts who operate in what Hogarth (2001) has called “kind environments,” where they can be expected to learn and improve, and counsels against giving too much deference to experts operating in what Hogarth calls “wicked environments,” where learning is hard. This emphasis on the dynamic nature of judgment and choice, and the prospect of learning and development, differs greatly from the static view found within BLE (see Klick and Mitchell 2006).

With respect to motivational assumptions, recently there have been efforts to move BLE toward a greater embrace of unconscious or implicit motivations (e.g., Levinson 2008; Levinson and Smith 2012). Adopting an implicit motivation framework, which gives precedence to associations formed over many years that operate unconsciously, could fundamentally alter BLE’s approach to the law and constraining behavior. While BLE rejects the L&E view of individuals as cold, calculating utility maximizers, BLE still assumes that intentional behavior based on a consideration of material incentives, albeit sometimes in an impatient and other-regarding way, will be the primary driver of behavior. An implicit motivation approach puts any stimuli with which positive or negative (or approach and avoidance) associations may have formed into play as a driver of behavior, making it very difficult to predict motivations and control behavior. Within this framework, the key question for BLE would become the extent to which the law can encourage individuals to shift into a more deliberative mode, overcome implicit preferences, and be driven by explicit cost-benefit calculations (see Jolls and Sunstein 2006). Presently, implicit motivation research does not provide clear guidance on when implicit, as opposed to explicit, attitudes and beliefs will drive behavior (see Oswald et al. 2013), and it would seem premature to adopt an implicit motivation framework given the many contentious issues surrounding this body of research (see, e.g., Mitchell and Tetlock, *in press*; Newell and Shanks 2014; Tetlock and Mitchell 2009).

An alternative motivational perspective that would give plans and intentions a larger role is the goal-based decision-making framework put forward by Krantz and Kunreuther (2007; see also Carlson et al. 2008). This framework assumes that



individuals attempt to achieve goals rather than maximize happiness or utility, and it explains as the product of multiple goal satisfaction rather than the product of K-T man's risk aversion or heuristic processing. When a decision is viewed narrowly, it may appear financially irrational (e.g., equivalent changes to utility may be treated differently depending on whether they are framed as potential losses or gains), but when extrafinancial goals are taken into account, the decision may appear rational.<sup>6</sup> That is not to say that all choices can be made to be rational within this broader perspective, because Krantz and Kunreuther argue that goals are context dependent and thus can lead to inconsistent and unwise choices through manipulation of the context and goals. Because this framework adopts a strong constructed-choice view, in which "true" preferences are "mythical" (Krantz and Kunreuther 2007, p. 160), the prescriptive focus falls on the process by which preferences are constructed. In this respect, it departs from the libertarian paternalism view that assumes that at least some choices for some groups of people reflect true preferences (see Sunstein and Thaler 2003; Mitchell 2005). Embracing Krantz and Kunreuther's perspective would force BLE to deal with the prescriptive problems presented by a strong constructed-preferences view.

Each of these perspectives, if adopted would lead to important clarifications or changes in how heuristics, biases, preferences, or irrationality are defined, in the causes of irrational behavior, or in the measures to be taken to improve judgment and decision-making. If a general reorientation along the lines of one of these alternative perspectives is hard to imagine (or some other perspective I failed to discuss, such as an approach that seeks to integrate affect into the decision process; e.g., Loewenstein et al. 2001), consider smaller-scale changes to particular findings and midlevel theories embraced by BLE scholars where those findings or theories have encountered challenges. For instance, Birnbaum has assembled considerable evidence casting doubt on the descriptive accuracy of prospect theory and cumulative prospect theory for risky choices, and on the supposed mechanism behind the reflection effect (e.g., Birnbaum 2006, 2008; Birnbaum et al. 2008), and Reiss (2004, 2005) explains why the proposition sometimes found in BLE articles that extrinsic rewards undermine intrinsic motivation is not warranted. Many examples could be offered, for the field of judgment and decision-making does not suffer from a lack of theoretical and empirical debate. The vibrancy of the field, and the importance of the topics found within it, guarantee that few noteworthy findings or theories will go unchallenged for long.

BLE's behavioral assumptions have been defended through comparison with the behavioral assumptions found in L&E, rather than in comparison to alternative behavioral propositions found within behavioral decision theory. Although it is not clear that an assumption of rational behavior is always such a bad choice descriptively and prescriptively (see Mitchell 2002b, 2003), certainly psychologists and behavioral

<sup>6</sup> The goal-based and expected-utility approaches to choice differ considerably, and the goal-based approach differs from prospect theory. For an extended discussion of these differences using the example of insurance, see Krantz and Kunreuther (2007).



economists have documented many reliable deviations from rationality. But which psychological theory or framework best predicts and explains those deviations—which version of bounded rationality is best—is a matter of considerable empirical dispute. Rather than select one all-purpose behavioral model, an alternative approach for BLE would be to take a problem-centered approach that replaces domain-general prescriptions with domain-specific solutions.

## 2 METHOD

If there is a method characteristic of the foundational (and many subsequent) BLE works, it is an analytical method and not an empirical method. Some BLE scholars do perform empirical research and adapt paradigms from behavioral decision theory to legal settings (e.g., Korobkin 1998; Rachlinski and Jourden 1998; see also chapter 5 by Engel in this volume on the growth of empiricism within BLE), but the great majority of BLE scholarship involves extrapolation from preexisting empirical research to some legal setting or task (see also Teichman 2011). The BLE scholar invokes a behavioral tendency supposedly revealed by the empirical studies, such as the status quo bias or base rate neglect, and applies it to some legal judgment or decision, such as judgments about future risks or past harms. The key consideration seems to be that the judgment or decision type under study be similar in kind to the legal judgment or decision of interest to the BLE scholar. The match between experimental conditions and applied conditions need not be great, and, as noted in the prior part, other behavioral tendencies and situational features that may complicate the picture are usually ignored (see, e.g., Mitchell 2002b, 2003).

BLE scholars rarely offer well-specified models of the incentive, informational, task, and other situational conditions under which irrational behavior is hypothesized to occur. This lack of specification means that BLE scholars rarely engage in a close comparison of the conditions present in the laboratory to the conditions in the setting of application. With respect to individual-level variables, BLE scholars typically treat K-T man as a fair representative of all humans, regardless of age, sex, education, and experience, sometimes claiming explicitly and at other times implying that most participants in the relevant empirical studies, or people in general, behave in a particular way.<sup>7</sup> That is very often not the case (Gigerenzer and Brighton 2009; Oreg and Bayazit 2009).

In some studies the majority of participants do exhibit irrational behavior. For example, in a meta-analysis of dictator game studies, Engel (2011) found that approximately 63% of all dictators offered more to the other player than a perfectly rational dictator

<sup>7</sup> BLE's analytical approach brings with it many unexamined assumptions and implications that I have criticized in detail elsewhere (Mitchell 2002a, 2003, in press). My overarching criticism is that BLE scholars often mistake internally valid findings for externally valid, generally applicable, practically significant findings. BLE scholars have too much faith that the experiments of

would, whereas approximately 36% exhibited rational behavior.<sup>8</sup> But those percentages only held in the aggregate: some settings and some groups exhibited very different patterns of behavior. Anonymous dictators gave less, older dictators gave more, and deserving recipients received more, among other variables that affected dictator giving. Engel (2011) describes the ultimatum game as “an exercise in exploring human *heterogeneity*” (p. 607, emphasis added), as “a powerful tool for [studying] the systematic variation of conditions that moderate sociality” (p. 608). An application of the ultimatum game results should take this systematic variation into account rather than reduce the results to the simple proposition that individuals have other-regarding preferences that may trump narrow self-interest (on the need for comparative empirical research pitting BLE and L&E predictions against one another, see Mitchell, in press).

BLE’s broad analytical approach permits BLE scholars tremendous prescriptive flexibility but also makes it likely that many prescriptions will fail to have their desired effects. The use of underspecified empirical assumptions protects K-T man in the face of failed predictions and prescriptions (if only X had been present, then Y would have occurred. . .) but interferes with the goal of injecting greater realism into the law: to advance the behavioral realism of our assumptions about how individuals interact with, interpret, and apply the law in particularized settings, we need to understand which conditions are necessary and which are sufficient for the production of which irrationalities. Or, more generally, we need to recognize the probabilistic relations among the relevant variables for particularized areas of the law and take those probabilities into account when formulating prescriptions.

There is a deeper problem, however, than the problem of faithful application of the varied experimental results: attempts to bring the outside world into the laboratory may not capture all of the variables that will be important once judgments and decisions are back outside the laboratory. The approach adopted by Kahneman and Tversky and utilized by many subsequent cognitive and social psychologists was to formulate a decision problem with a rational solution derived from the postulates of rational choice theory and an irrational solution derived from a psychological theory. The context supplied in the judgment or decision problem itself is key (e.g., how choice options are framed, whether nondiagnostic or pseudodiagnostic information is provided), not the larger environment in which the judgment or situation occurs. These studies were designed to test whether people deviate from rationality in the ways predicted by psychological theories, not to examine the generality of the observed phenomena. Edwards (1983) notes

psychology and behavioral economics have identified robust, general tendencies instead of localized demonstrations of deviations from rational choice theory. I focus here on the broad contours of this approach and some alternatives that might produce more reliable prescriptions tailored to specific problems.

<sup>8</sup> That approximately one-third of the more than 20,000 dictators studied by Engel exhibited rational behavior is not itself a trivial finding. According to Engel (2011), this finding indicates that “generosity is not a human universal” (p. 607). Even among those who gave more than was rational, they did not exhibit perfect generosity in terms of an even split; they kept on average about 8% more of the pie for themselves.

with regret that the approach taken in his pioneering research served as a precursor to Kahneman and Tversky's method. Edwards's experiments focused on finding errors in probabilistic reasoning and were an outgrowth of error-focused psychological research on memory and perception, where errors were used to understand the functioning and limits of human sensory abilities. In the sensory research, the tasks and perceptual abilities were seen as so basic that any stimuli could be used and any individual's mind was assumed to be representative of normal adult minds (including the researcher's own mind, which explains why introspection was often used in this research). Use of this approach in memory studies turned out to be particularly unfortunate, given what we now know about the importance of individual differences in cognitive ability and about features of the stimuli and task to remembering.

Because BLE searches for empirical regularities that can serve as the basis for legal prescriptions, rather than tests theories, the external validity of the studies that BLE scholars rely on should be a primary concern (see Schram 2005). Unfortunately, the findings of many of these studies probably do not generalize to the applied domains of interest to BLE scholars. In a recent study of effects found in the laboratory that were subjected to testing in the field, I found that the effects from social psychology laboratories (the source of many biases invoked by BLE scholars) often did not replicate in the field, and in many instances the direction of the effect was different in the field than that observed in the laboratory (Mitchell 2012). This lack of external validity can be attributed to the generally small size of effects studied by social psychologists (small effects are less likely to replicate),<sup>9</sup> but also to the purpose behind the studies being theory testing rather than establishment of empirical regularities (or demonstrations of effects rather than understanding of the prevalence of behaviors). There is thus a mismatch between what behavioral decision theory does and what BLE needs.

One alternative methodological approach would be based on what Santos (2007) calls technological experiments within experimental economics. Technological experiments, such as those undertaken by Plott (1997) and Roth (1984) on the design of markets, seek to identify the conditions under which an institution can achieve a particular goal where the institution has considerable control over the information (and often choices) available and how that information is used. Field experiments undertaken to test Thaler's SMarT savings program (Benartzi and Thaler 2004) are a good example of using experiments as tools of institutional engineering: begin with an understanding of how individuals are likely to behave with respect to well-identified decisions, propose a situation-specific intervention to achieve an institutionally defined goal with respect to those decisions, and determine whether the intervention produces the desired effects.

<sup>9</sup> Some of the most prominent effects identified by Kahneman and Tversky have turned out to be quite small. For instance, Gallagher and Updegraff (2012) synthesized studies examining the impact of gain versus loss framing on decisions to detect or prevent poor health behaviors and found no reliable effect for framing with respect to detection behaviors, and found a small effect for framing on prevention behavior, which is consistent with Kühberger's (1998) earlier meta-analysis that found framing effects to be "only small to moderate size" (p. 42).

This technological approach may be generalized by emphasizing mundane realism and problem-solving, akin to what Roth (2002) calls “design economics.” In this approach, a problem is identified, and then that problem is subjected to high-fidelity experimental and field studies designed to identify the cause of the problem and possible solutions that get refined over time. The key is that the problem be faithfully simulated in the lab, to ensure that important task and environmental factors are captured, and be studied in the wild, to verify the laboratory results. This approach is popular in industrial-organizational psychology and leads to results with remarkably good external validity (Mitchell 2012). Fischhoff (1996) advocates a similar approach that involves gradual complication of the empirical studies to make sure that important variables are captured and that emphasizes ensuring that the researcher understands the judgment or decision problem from the perspective of the targets.

The foregoing alternatives would require greater involvement in empirical research by BLE scholars, many of whom do not have empirical training or the resources to conduct such research. Some may pair with social scientists to conduct original research; many will not. Therefore, a more realistic alternative would involve changes to the analytical approach presently favored by BLE scholars, and my candidate for the most feasible and most important change to this method involves taking the empirical research as it exists but seeking to divine in it the necessary and sufficient conditions for the domain-specific appearance of rational versus irrational behaviors. This change would begin with the scholar unpacking the usually implicit assumptions made about necessary and sufficient conditions: in the research literature relied on, do individuals vary by age, sex, or other identifiable characteristics in their risk aversions or whatever behavioral disposition is of interest, does the proposed behavior vary by type of good or contract amount, does experience moderate the effect, and so on, and which of these conditions are necessary and which are sufficient? Then the scholar would ask whether these conditions approximate the conditions that hold in the legal setting of interest.

An important byproduct of this process is the identification of conditions under which the K-T man model, or whatever model is chosen, is likely not to hold. As Roth (2002) discusses, with applied work, unlike basic research, complications cannot be shoved under the rug, so a sense of the complications likely to arise and their magnitude is needed. BLE scholars often have a good sense of the unwanted outcomes or irrational behaviors they hope to help overcome through their extrapolations from behavioral decision theory research, but they may have a poor sense of the complications likely to arise or may simply not be focused on the complications.

This proposed process may appear foolhardy given how complex some legal problems can be and how underdeveloped the empirical research can be,<sup>10</sup> and that is

<sup>10</sup> Edwards (1983) estimated that there were at least 3,456 combinations of variables that could be important to judgments and decisions but that only 32 variable combinations had been studied as of the time he was writing. That figure has probably not increased tremendously in the succeeding thirty years.

precisely the point: if it is difficult to specify the behaviorally important features of the legal setting or to draw from the existing literature a set of necessary and sufficient conditions for a phenomenon that correspond to these important features, then that difficulty should encourage prescriptive caution and highlight the need for additional research describing empirical regularities within the legal setting or identifying necessary and sufficient conditions within the psychological or behavioral economic literature (for a discussion of how to proceed in the face of empirical uncertainty, see Mitchell 2002a, 2004).

In this alternative world, BLE scholars would push psychologists and behavioral economists to produce more helpful empirical research by identifying gaps in our understanding and encouraging greater separation between the conditions empirically determined to be necessary and sufficient for a theory to hold and the auxiliary assumptions that are neither necessary nor sufficient but have simply been chosen to implement a test of a theory (e.g., are the particular stimuli chosen seen as crucial or fungible?). This alternative world would still leave us a good distance from reliable prescriptions. As Fischhoff (1996) puts it, “Creating workable interventions requires detailed design work, followed by rigorous empirical evaluation” (p. 244). Or as Roth (2002) puts it, “design involves a responsibility for detail; this creates a need to deal with complications” (p. 1342). But it would be a step in the right direction if the goal is particularized prescriptions based on realistic behavioral assumptions. If the goal is just to do battle with the L&E over the basic assumptions about human nature that should be used broadly to guide legal prescriptions, then concerns still remain about the particular model chosen by BLE, namely, K-T man instead of one of its competitors. That human nature project is important and may well be worth fighting scientifically—as Fischhoff (1996) discusses, his research on adolescent risk taking may have had its greatest real-world impact not through any specific policy but by altering basic views about adolescent competence that can translate into greater or lesser autonomy—but I am treating the BLE project as an attempt to move the law toward evidence-based prescriptions in order to promote more effective laws, whatever the underlying goals may be (see, e.g., Sunstein 2011).

Whatever approach is taken by BLE scholars going forward, at least two rules should constrain BLE prescriptions. First, the focus should be on discovering, and basing prescriptions on, reliable patterns of behavior across multiple studies. Findings based on a handful of experimental studies should be given no weight in BLE’s prescriptions. Surprising results may provide grounds for questioning L&E assumptions but should not be the basis for prescriptions until their robustness is demonstrated: social science is littered with counterintuitive findings that failed to survive scrutiny. Second, before embracing a behavioral finding as its own, BLE scholars should insist that the finding be replicated with samples from their target populations in settings, using tasks and stimuli, that approximate those in the target domain (cf. Marewski, Schooler, and Gigerenzer 2010). That insistence may mean that BLE cannot match every prescription made by L&E as it waits on empirical evidence, but it also means that BLE will become a more trustworthy guide to prescriptions. These rules might usefully be applied to

L&E as well, leading to the rejection of many L&E prescriptions as well, and leaving policymakers with competing claims about how to proceed with little empirical evidence to guide them. That state of affairs would, in my opinion, be preferable because it would force decisions to be made on value-driven grounds (e.g., would an error from following the BLE prescription be better than an error from following the L&E prescription?) rather than on mistaken belief about the state of our empirical knowledge. And it might lead to better problem-focused field research in the drive for evidence-based regulations (see Sunstein 2011).

### 3 UTILITY MAXIMIZATION AND THE LAW AS A PRICE

---

The law per se plays an oddly limited role within BLE.<sup>11</sup> BLE scholars accept L&E's consequentialism and orientation toward law as a price that can be adjusted up or down to achieve various effects (see Posner 2001, 2010), with BLE scholars emphasizing how biases in beliefs and in the evaluation of options can cause inconsistencies in how actors react to legal incentives *ex ante* or how they impose rewards and penalties after the fact (for a critique of this framework in the criminal law context, see chapter 22 by Harel in this volume). BLE scholars criticize, that is, the descriptive accuracy of utility maximization but not its normative status (Huigens 2001).<sup>12</sup> Many of BLE's prescriptions are designed to help people make the utility-maximizing choices they would presumably make were they rational or were they doing as some authority believes they should (e.g., by setting particular default terms in contracts or imposing cooling-off periods to

<sup>11</sup> For instance, L&E scholars devote considerable attention to the form a law takes—broadly, a rule versus a standard (e.g., Johnston 1995; Kaplow 1992)—but to my knowledge BLE scholars have not given equal attention to this question from a behavioral perspective (for some preliminary thoughts on the topic, see Hirsch and Mitchell 2008). Where people may not realize that the law provides a structure, stricture, or informational baseline for transactions, as with usury limits or required disclosures on loans and consumer goods, BLE's lack of attention to how people understand, internalize, and react to the law or legal mandates is understandable. However, even in situations where the salience of the law's role is low, the manner in which information is presented and the source of that information may hold greater importance than just creating an endowment or setting returns as losses or gains relative to some reference point. Sher and McKenzie's information leakage theory (2006, 2011) emphasizes that in most framing situations the frames chosen may be logically, but not informationally, equivalent. Thus, setting a default rule may communicate what choice the authority believes is appropriate (e.g., McKenzie et al. 2006), which may lead to greater acquiescence or resistance depending on one's views of the authority or source (see, e.g., Druckman 2001, 2004; Druckman and Chong, *in press*). Along somewhat similar lines, Wilkinson-Ryan (2010) found that including a liquidated damages clause in a contract signaled to parties (in the role of hypothetical contractors) that an efficient breach would be more acceptable than when damages were not provided for in the contract.

<sup>12</sup> There is a fundamental tension between the use of propositions from expected utility theory as normative ideals and the use of behavioral evidence to set legal prescriptions because utility theory leaves it to the individual to define utility: we determine rationality not through reference to some



make up for willpower deficits), and some prescriptions are aimed at preventing firms from exploiting irrationalities (e.g., by restricting certain choices for some groups of people or imposing liability for some kinds of behavior). Other BLE prescriptions implement a broader conception of utility that incorporates other-regarding preferences but still operate within an expected utility framework.

Oftentimes people depart from rational choice predictions not because they are irrational but because they reject efficiency or utility maximization as the proper goal. Baron has studied several nonconsequentialist rules that people employ to make decisions with legal or social welfare implications (see, e.g., Baron 1994, 1998; chapter 3 in this volume). Sunstein (2005) shares Baron's view that many moral and legal judgments are based on "moral heuristics" that can lead people away from normative ideals. Sunstein defends a weakly consequentialist normative framework, but he argues that moral heuristics may lead to deviations from whatever normative framework one adopts. Baron and Sunstein reject the view popular among some philosophers that intuitions are a reliable guide to value judgments and discovery of "moral facts."

I do not doubt that a good number of legal and moral judgments are the product of intuitive judgments that people would reconsider under some conditions, and I am skeptical of treating intuitions as reliable guides to moral facts or values (setting aside the difficult question of how to separate post hoc rationalizations from reason-giving arguments; see Mitchell and Tetlock 2009). But some nonconsequentialist moral and legal judgments and decisions would appear to be hard to characterize as unreasoned and mistaken from an internal point of view, and probably reflect either an intentional rejection of efficiency as the right goal or a trade-off between efficiency and another value (cf. Mitchell et al. 1993, 2003). In the domain of torts, for instance, Tetlock and I found that a corrective justice norm, more than deterrence or distributive justice concerns, drove judgments about liability following both intentional and negligent torts (Mitchell and Tetlock 2006).

The epistemic status of intuitions, which has assumed a major role in recent psychological studies of morality (Haidt and Kesibir 2010), presents important questions about the underpinnings of legal sanctions, the directions the law should take, and the likely effectiveness of laws that depart from robust and widespread intuitions (see, e.g., Robinson and Kurzban 2007; Jones and Kurzban 2010). Sunstein's article and Baron's extensive work could serve as the basis for an ambitious research program examining

objective good but by examining the coherence and consistency of an individual's choices across options. If we cannot rely on revealed or stated preferences to set prescriptions because of concerns about systematic biases or instability of preferences, then prescriptions will either need to be in the form of procedural reforms aimed at overcoming biases or based on some objective ideal of welfare that is imposed on people (i.e., substantive paternalism of some form). Libertarian paternalism is one attempt to overcome this difficulty (by taking a procedural approach), but it presents its own problems and ultimately requires externally imposed goals (see Mitchell 2005; see also Korobkin 2009; Rebonato 2012). Where a public or private institution with power over individuals specifies the directions in which those individuals should be pointed, this tension dissolves, but in many domains individuals are left to their own devices to set goals and the law ostensibly serves only to structure transactions or regulate self-determined transactions.



the prevalence, sources, robustness, and implications of Sunstein's proposed moral heuristics, but BLE scholars have shown little interest in taking up this research program on their own or in collaboration with social scientists (for examples of exceptions, see Depoorter and Tontrup 2012; Wilkinson-Ryan 2011). Questions about why people sometimes reject distributional efficiency and transactional utility maximization, or trade efficiency off against other values, raise some of the thorniest and most interesting questions for legal and political systems. BLE scholars are particularly well suited to address these questions because, through their legal expertise, they possess sophisticated understanding of the arguments that can be made, and tested, in support of deontological as opposed to consequentialist norms.

BLE should take a more behavioral approach to norm selection and legal interpretation if for no other reason than to improve its prescriptions. It may be that people penalize efficient risk-taking because they fall prey, say, to a "normality bias" (Prentice and Koehler 2003), or some other bias that leads to error. Or it may be that people worry about the larger social or motivational effects of embracing a particular normative rule or want to re-establish the status of group members who suffered harm (e.g., Tetlock 2002). A host of interesting questions arise if we ask whether a deviation from consequentialism was caused by an unwanted judgmental bias or by rejection of consequentialism, rather than assume the former to be the case. When and why do people reject the notion of efficient breach (Wilkinson-Ryan and Baron 2009)? Why are some people insufficiently sensitive to the probability of detection in tax compliance behavior (cf. Baron and Ritov 2009)? And so on. In some situations, many individuals would surely endorse the goal of utility maximization and welcome the law as a means of facilitating efficient transactions, either through *ex ante* restrictions (e.g., antitrust laws) or *post hoc* review, where bounds on rationality get in the way of efficiency. But in other situations, L&E and BLE will confront the more fundamental problem of convincing individuals that a consequentialist approach is the right approach. Many behavioral findings are likely to be relevant regardless of the normative theory embraced, but they may have different implications for different normative theories (see, for example, the discussion of the normative implications loss aversion research in Zamir 2012).

## 4 CONCLUSION

---

Please choose one of the following options:

- (a) Many of the insights found in BLE are clever and original but are based on a fairly narrow set of psychological assumptions and involve questionable applications of psychological research.
- (b) Until significant changes are made to move BLE toward empirically tested, domain-specific prescriptions, there is little reason to believe that BLE's prescriptions will be more effective than those offered by L&E.

- (c) Employing better specified models would be an important first step toward a truly behavioral approach to law of the kind envisioned by early BLE articles.
- (d) None of the changes imagined in this chapter, especially the methodological ones, is likely to be embraced by BLE scholars.
- (e) All of the above.
- (f) None of the above.

My choice was (e). I hope your choice was not (f), because if it was, then you made a choice that rejects option (d), which you probably did not mean to reject. Did you do that because you are incapable of simple logic, because you relied on a simple heuristic that usually produces good results (such as “always disagree with this author”), because you read the options to mean something different than I intended, or because you were momentarily inattentive? You probably should have just skipped the question given that answering it is of no consequence—that would have been the efficient thing to do—but of course that was not an explicit option, and maybe you were just trying to be a good participant and select the best option from a set of unattractive options. Or maybe you were trying to be nice to the author, even though I will probably never ask you whether you even answered the question, much less what answer you gave. If you did select (f) with the intention of rejecting option (d), then what does that say about your understanding of the world and the influence of book chapters in it, particularly one written by this author? In that case, we will count you as logical, but you probably do need some help. Whatever your choice, I am confident that I could find in BLE guidance on how to present this problem in a way that would allow me to demonstrate some supposed irrationality on your part, no matter who you are or what the problem involves. If you doubt my ability to do that, then perhaps you should take my concerns about BLE more seriously. Or, if you get the point of this silly little demonstration, then perhaps you should take my concerns about BLE more seriously.

## REFERENCES

- Baker, J. A. 2011. And the winner is: How principles of cognitive science resolve the plain language debate. *UMKC Law Review* 80:287–305.
- Baron, J. 1994. Nonconsequentialist decisions (including comments and reply). *Behavioral and Brain Sciences* 17:1–42.
- Baron, J. 1998. *Judgment misguided: Intuition and error in public decision making*. New York: Oxford University Press.
- Baron, J. 2012. The point of normative models in judgment and decision making. *Frontiers in Psychology* 3: Article 577, 1–3.
- Baron, J. and Ritov, I. 2009. The role of probability of detection in judgments of punishment. *Journal of Legal Analysis* 1: 553–90.
- Benartzi, S., and Thaler, R. H. 2004. Save more tomorrow: Using behavioral economics to increase employee savings. *Journal of Political Economy* 112: S164–87.
- Birnbaum, M. H. 2006. Evidence against prospect theories in gambles with positive, negative, and mixed consequences. *Journal of Economic Psychology* 27: 737–61.

- Birnbaum, M. H. 2008. New paradoxes of risky decision making. *Psychological Review* 115: 463–501.
- Birnbaum, M. H., Johnson, K., Longbottom, J. L. 2008. Tests of cumulative prospect theory with graphical displays of probability. *Judgment and Decision Making* 3: 528–46.
- Brunswik, E. 1956: *Perception and the representative design of psychological experiments*. 2nd ed. Berkeley: University of California Press.
- Carlson, K. A., Janiszewski, C., Keeney, R. L., Krantz, D. H., Kunreuther, H. C., Luce, M. F., Russo, J. E., van Osselaer, S. M. J., and von Winterfeldt, D. 2008. A theoretical framework for goal-based choice and for prescriptive analysis. *Marketing Letters* 19: 241–54.
- Dawes, R. M. 1985. Behavioral decision making and judgment. In G. Lindzey and E. Aronson (Eds.), *Handbook of social psychology*, 3rd ed., vol. 2, pp. 176–97. New York: Random House.
- Denrell, J., and Le Mens, G. 2012. Social judgments from adaptive samples. In J. I. Krueger (Ed.), *Social judgment and decision-making*, pp. 151–69. New York: Psychology Press.
- Depoorter, B., and Tontrup, S. 2012. How law frames moral intuitions: The expressive effect of specific performance. *Arizona Law Review* 54: 673–717.
- Dunwoody, P. T. 2009. Theories of truth as assessment criteria in judgment and decision making. *Judgment and Decision Making* 4: 116–25.
- Edwards, W. 1954. The theory of decision making. *Psychological Bulletin* 41: 380–417.
- Edwards, W. 1961. Behavioral decision theory. *Annual Review of Psychology* 12: 473–98.
- Edwards, W. 1983. Human cognitive capabilities, representativeness, and ground rules for research. In P. C. Humphreys, O. Svenson, and A. Vari (Eds.), *Analyzing and aiding decision processes*, pp. 507–13. Amsterdam: North-Holland.
- Engel, C. 2011. Dictator games: A meta study. *Experimental Economics* 14: 583–610.
- Fiedler, K. 2000. Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review* 107: 659–76.
- Fiedler, K., and Juslin, P. 2006. Taking the interface between mind and environment seriously. In K. Fiedler and P. Juslin (Eds.), *Information sampling and adaptive cognition*, pp. 3–29. New York: Cambridge University Press.
- Fischhoff, B. 1996. The real world: What good is it? *Organizational Behavior and Human Decision Processes* 65: 232–48.
- Gallagher, K. M., and Updegraff, J. A. 2012. Health message framing effects on attitudes, intentions, and behavior: A meta-analytic review. *Annals of Behavioral Medicine* 43: 101–16.
- Gigerenzer, G. 2000. *Adaptive thinking: Rationality in the real world*. Oxford: Oxford University Press.
- Gigerenzer, G. 2006. Heuristics. In G. Gigerenzer and C. Engel (Eds.), *Heuristics and the law*, pp. 17–44. Cambridge, MA: MIT Press.
- Gigerenzer, G., and Brighton, H. 2009. Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science* 1: 107–43.
- Gilovich, T., Griffin, D., and Kahneman, D. (Eds.). 2002. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.
- Goldstein, W. M. 2004. Social judgment theory: Applying and extending Brunswik's probabilistic functionalism. In Derek J. Koehler and Nigel Harvey (Eds.), *Blackwell handbook of judgment and decision making*, pp. 37–61. Oxford: Blackwell.
- Guthrie, C., Rachlinski, J. J., and Wistrich, A. J. 2001. Inside the judicial mind. *Cornell Law Review* 86: 777–830.
- Haidt, J., and Kesebir, S. 2010. Morality. In S. Fiske, D. T. Gilbert, and G. Lindzey (Eds.), *Handbook of social psychology*, 5th ed., vol. 2, pp. 797–832. Hoboken, NJ: John Wiley and Sons.

- Hans, V. P., and Reyna, V. F. 2011. To dollars from sense: Qualitative to quantitative translation in jury damage awards. *Journal of Empirical Legal Studies* 8: 120–47.
- Hanson, J. D., and Kysar, D. A. 1999a. Taking behavioralism seriously: Some evidence of market manipulation. *Harvard Law Review* 112: 1420–571.
- Hanson, J. D., and Kysar, D. A. 1999b. Taking behavioralism seriously: The problem of market manipulation. *New York University Law Review* 74: 101–217.
- Hastie, R. 2001. Problems for judgment and decision making. *Annual Review of Psychology* 52: 653–83.
- Hilbert, M. 2012. Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin* 138: 211–37.
- Hirsch, A. J., and Mitchell, G. 2008. Law and proximity. *University of Illinois Law Review* 2008: 557–98.
- Hogarth, R. M. 2001. *Educating intuition*. Chicago: University of Chicago Press.
- Huigens, K. 2001. Law, economics, and the skeleton of value fallacy. *California Law Review* 89: 969–96.
- Jolls, C., and Sunstein, C. R., 2006. The law of implicit bias. *California Law Review* 94: 969–96.
- Jolls, C., Sunstein, C. R., and Thaler, R. 1998. A behavioral approach to law and economics. *Stanford Law Review* 50: 1471–550.
- Johnston, J. S. 1995. Bargaining under rules versus standards. *Journal of Law, Economics and Organization* 11: 256–81.
- Jones, O. D., and Kurzban, R. 2010. Intuitions of punishment. *University of Chicago Law Review* 77: 1633–41.
- Kahneman, D. 2011. *Thinking, fast and slow*. New York: Farrar Straus and Giroux.
- Kahneman, D., and Frederick, S. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. W. Griffin, and D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment*, pp. 49–81. New York: Cambridge University Press.
- Kahneman, D., and Frederick, S. 2005. A model of heuristic judgment. In K. J. Holyoak and R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning*, pp. 267–93. Cambridge: Cambridge University Press.
- Kahneman, D., and Klein, G. 2009. Conditions for intuitive expertise: A failure to disagree. *American Psychologist* 64: 515–26.
- Kahneman, D., Slovic, P., and Tversky, A. 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kahneman, D., and Tversky, A. (Eds.). 2000. *Choices, values, and frames*. Cambridge: Cambridge University Press.
- Kaplow, L. 1992. Rules versus standards: An economic analysis. *Duke Law Journal* 42: 557–629.
- Kelman, M. 2011. *The heuristics debate*. Oxford: Oxford University Press.
- Klein, G. 1998. *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.
- Klick, J., and Mitchell, G. 2006. Government regulation of irrationality: Moral and cognitive hazards. *Minnesota Law Review* 90: 1620–63.
- Korobkin, R. B. 1998. The status quo bias and contract default rules. *Cornell Law Review* 83: 608–87.
- Korobkin, R. B. 2009. Libertarian welfarism. *California Law Review* 97: 1651–85.
- Korobkin, R. B. 2011. What comes after victory for behavioral law and economics? *University of Illinois Law Review* 2011: 1653–74.
- Korobkin, R. B., and Guthrie, C. 1994. Psychological barriers to litigation settlement: An experimental approach. *Michigan Law Review* 93: 107–92.

- Korobkin, R. B., and Guthrie, C. 1997. Psychology, economics, and settlement: A new look at the role of the lawyer. *Texas Law Review* 76: 77–141.
- Korobkin, R. B., and Ulen, T. S. 2000. Law and behavioral science: Removing the rationality assumption from law and economics. *California Law Review* 88: 1051–144.
- Krantz, D. H., and Kunreuther, H. C. 2007. Goals and plans in decision making. *Judgment and Decision Making* 2: 137–68.
- Krueger, J. I. 2012. The (ir)rationality project in social psychology: A review and assessment. In J. I. Krueger (Ed.), *Social judgment and decision-making*, pp. 59–75. New York: Psychology Press.
- Krueger, J. I., and Funder, D. C. 2004. Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social behavior and cognition (with comments and reply). *Behavioral and Brain Sciences* 27: 313–76.
- Kühberger, A. 1998. The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes* 75: 23–55.
- Langevoort, D. C. 1992. Theories, assumptions, and securities regulation: Market efficiency revisited. *University of Pennsylvania Law Review* 140: 851–920.
- Langevoort, D. C. 1995. Ego, human behavior, and law. *Virginia Law Review* 81: 853–86.
- Levinson, J. D. 2008. Culture, cognitions and legal decision-making. In R. Sorrentino and S. Yamaguchi (Eds.), *Handbook of motivation and cognition across cultures*, pp. 423–41. San Diego, CA: Elsevier Press.
- Levinson, J. D., and Smith, R. J. 2012. *Implicit racial bias across the law*. Cambridge: Cambridge University Press.
- Loewenstein, G., Weber, E., Hsee, C., and Welch, N. 2001. Risks as feelings. *Psychological Bulletin* 127: 267–86.
- Marewski, J. N., Schooler, L. J., and Gigerenzer, G. 2010. Five principles for studying people's use of heuristics. *Acta Psychologica Sinica* 42: 72–87.
- McFadden, D. L. 1999. Rationality for economists. *Journal of Risk and Uncertainty* 19: 73–105.
- Mellers, B. A., Hertwig, R., and Kahneman, D. 2001. Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science* 12: 269–75.
- Mitchell, G. 2002a. Taking behavioralism too seriously? The unwarranted pessimism of the new behavioral analysis of law. *William and Mary Law Review* 43: 1907–2021.
- Mitchell, G. 2002b. Why law and economics' perfect rationality should not be traded for behavioral law and economics' equal incompetence. *Georgetown Law Journal* 91: 67–167.
- Mitchell, G. 2003. Tendencies versus boundaries: Levels of generality in behavioral law and economics. *Vanderbilt Law Review* 56: 1781–812.
- Mitchell, G. 2004. Case studies, counterfactuals, and causal explanations. *University of Pennsylvania Law Review* 152: 1517–608.
- Mitchell, G. 2005. Libertarian paternalism is an oxymoron. *Northwestern University Law Review* 99: 1245–77.
- Mitchell, G. 2009. Second thoughts. *McGeorge Law Review* 40: 687–722.
- Mitchell, G. 2012. Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science* 7: 109–17.
- Mitchell, G. In press. The price of abstraction. In J. C. Teitelbaum and K. Zeiler (Eds.), *Research handbook on behavioral law and economics*.
- Mitchell, G., and Tetlock, P. E. 2006. An empirical inquiry into the relation of corrective justice to distributive justice. *Journal of Empirical Legal Studies* 3: 421–66.

- Mitchell, G., and Tetlock, P. E. 2009. Disentangling reasons and rationalizations: Exploring fairness in hypothetical societies. In J. Jost, A. C. Kay, and H. Thorisdottir (Eds.), *Social and psychological bases of ideology and system justification*, pp. 126–57. New York: Oxford University Press.
- Mitchell, G., and Tetlock, P. E. In press. Implicit attitude measures. In R. A. Scott and S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences*. Thousand Oaks, CA: Sage Publications.
- Mitchell, G., Tetlock, P. E., Mellers, B., and Ordóñez, L. 1993. Judgments of social justice: Compromises between equality and efficiency. *Journal of Personality and Social Psychology* 65: 629–39.
- Mitchell, G., Tetlock, P. E., Newman, D. G., and Lerner, J. S. 2003. Experiments behind the veil: Structural influences on judgments of social justice. *Political Psychology* 24: 519–47.
- Newell, B. R. and Shanks, D. R. 2014. Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences* 37: 1-19.
- Oreg, S., and Bayazit M. 2009. Prone to bias: Towards a theory of individual differences in bias manifestation. *Review of General Psychology* 13: 175–93.
- Osborne, G., and Thaler, R. Jan. 28, 2010. We can make you behave. *The Guardian*. Available at <http://www.guardian.co.uk/commentisfree/2010/jan/28/we-can-make-you-behave>.
- Oswald, F., Mitchell, G., Blanton, H., Jaccard, J., and Tetlock, P. E. 2013. Predicting ethnic and racial discrimination: A new meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology* 105: 171–192.
- Paredes, T. A. 2003. Blinded by the light: Information overload and its consequences for securities regulation. *Washington University Law Quarterly* 81: 417–85.
- Pliske, R., and Klein, G. 2003. The naturalistic decision-making perspective. In S. Schneider and J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research*, pp. 559–85. New York: Cambridge University Press.
- Plott, C. 1997. Laboratory experimental testbeds: Applications to the PCS auction. *Journal of Economics and Management Strategy* 6: 605–38.
- Posner, R. A. 2001. *Frontiers of legal theory*. Cambridge, MA: Harvard University Press.
- Posner, R. A. 2010. *Economic analysis of law*. 8th ed. New York: Aspen Publishers.
- Prentice, R. A. 2003. Chicago man, K-T man, and the future of behavioral law and economics. *Vanderbilt Law Review* 56: 1663–777.
- Prentice, R. A. and Koehler, J. J. 2003. A normality bias in legal decision making. *Cornell Law Review* 88: 583–650.
- Rachlinski, J. J. 1996. Gains, losses, and the psychology of litigation. *Southern California Law Review* 70: 113–85.
- Rachlinski, J. J. 1998. A positive psychological theory of judging in hindsight. *University of Chicago Law Review* 65: 571–625.
- Rachlinski, J. J. 2000. The new law and psychology: A reply to critics, skeptics, and cautious supporters. *Cornell Law Review* 85: 739–66.
- Rachlinski, J. J. 2011. The psychological foundations of behavioral law and economics. *University of Illinois Law Review* 2011: 1676–96.
- Rachlinski, J. J. 2012. Selling heuristics. *Alabama Law Review* 64: 389–415.
- Rachlinski, J. J., and Jourden, F. 1998. Remedies and the psychology of ownership. *Vanderbilt Law Review* 51: 1541–82.
- Rachlinski, J. J., Wistrich, A. J., and Guthrie, C. 2006. Inside the bankruptcy judge's mind. *Boston University Law Review* 86: 1227–65.



- Rebonato, R. 2012. *Taking liberties: A critical examination of libertarian paternalism*. New York: Palgrave Macmillan.
- Reiss, S. 2004. Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. *Review of General Psychology* 8: 179–93.
- Reiss, S. 2005. Extrinsic and intrinsic motivation at 30: Unresolved scientific issues. *Behavioral Analyst* 28: 1–14.
- Reyna, V. F. 2012. A new intuitionism: Meaning, memory, and development of fuzzy-trace theory. *Judgment and Decision Making* 7: 332–59.
- Reyna, V. F., Lloyd, F. J., and Brainerd, C. J. 2003. Memory, development, and rationality: An integrative theory of judgment and decision-making. In S. Schneider and J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research*, pp. 201–45. New York: Cambridge University Press.
- Richardson, L. S., and Gofff, P. A. 2012. Self-defense and the suspicion heuristic. *Iowa Law Review* 98: 293–336.
- Robinson, P. H., and Kurzban, R. 2007. Concordance and conflict in intuitions of justice. *Minnesota Law Review* 91: 1829–1907.
- Roth, A. E. 1984. The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy* 92: 991–1016.
- Roth, A. E. 2002. The economist as engineer: Game theory, experimental economics and computation as tools of design economics. *Econometrica* 70: 1341–78.
- Santos, A. C. 2007. The “materials” of experimental economics: Technological versus behavioral experiments. *Journal of Economic Methodology* 14: 311–37.
- Schram, A. 2005. Artificiality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology* 12: 225–38.
- Sedlmeier, P. 1999. *Improving statistical reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Sher, S., and McKenzie, C. R. M. 2006. Information leakage from logically equivalent frames. *Cognition* 101: 467–94.
- Sher, S., and McKenzie, C. R. M. 2011. Levels of information: A framing hierarchy. In G. Keren (Ed.), *Perspectives on framing*, pp. 35–63. New York: Psychology Press.
- Stanovich, K. E. 1999. *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Stanovich, K. E., and West, R. F. 2008. On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology* 94: 672–95.
- Stone, E. R., Yates, J. F., and Parker, A. M. 1994. Risk communication: Absolute versus relative expressions of low-probability risks. *Organizational Behavior and Human Decision Processes* 60: 387–408.
- Sunstein, C. R. 1998. Selective fatalism. *Journal of Legal Studies* 27: 799–823.
- Sunstein, C. R. 2005. Moral heuristics (with comments and reply). *Behavioral and Brain Sciences* 28: 531–73.
- Sunstein, C. R. 2011. Empirically informed regulation. *University of Chicago Law Review* 78: 1349–429.
- Sunstein, C. R., Kahneman, D., and Schkade, D. 1998. Assessing punitive damages (with notes on cognition and valuation in law). *Yale Law Journal* 107: 2071–153.
- Sunstein, C. R. and Thaler, R. (2003). Libertarian paternalism is not an oxymoron. *University of Chicago Law Review* 70: 1159–202.
- Teichman, D. 2011. The optimism bias of the behavioral analysis of crime control. *University of Illinois Law Review* 2011: 1697–712.



- Tetlock, P. E. 2002. Social-functional frameworks for judgment and choice: The intuitive politician, theologian, and prosecutor. *Psychological Review* 109: 451–72.
- Tetlock, P. E. 2005. *Expert political judgment: How good is it? How can we know?* Princeton, NJ: Princeton University Press.
- Tetlock, P. E., and Mitchell, G. 2009. Implicit bias and accountability systems: What must organizations do to prevent discrimination? *Research in Organizational Behavior* 28: 3–38 (Barry Staw and Arthur Brief, eds.).
- Wilkinson-Ryan, T. 2010. Do liquidated damages encourage breach? A psychological experiment. *Michigan Law Review* 108: 633–71.
- Wilkinson-Ryan, T. 2011. Breaching the mortgage contract: The behavioral economics of strategic default. *Vanderbilt Law Review* 64: 1547–83.
- Wilkinson-Ryan, T., and Baron, J. 2009. Moral judgment and moral heuristics in breach of contract. *Journal of Empirical Legal Studies* 6: 405–23.
- Wistrich, A. J., Guthrie, C., and Rachlinski, J. J. 2005. Can judges ignore inadmissible information? The difficulty of deliberately disregarding. *University of Pennsylvania Law Review* 153: 1251–345.
- Zamir, E. (2012). Loss aversion and the law. *Vanderbilt Law Review* 65: 829–94.



PART III

---

**BEHAVIORAL  
ECONOMICS AND  
THE LAW**

*Specific Behavioral Phenomena*

---



## CHAPTER 8

---

# LAW AND PROSOCIAL BEHAVIOR

---

LYNN A. STOUT

## 1 INTRODUCTION

---

A standard convention in the economic analysis of law for many years has been the so-called *homo economicus* model of human behavior. According to the *homo economicus* model, human beings are rational and selfish actors. This means that they rationally pursue their own self-interest and “utility,” which, as discussed below, is conventionally conceived of as maximizing their own material wealth.

### 1.1 History of the *Homo Economicus* Assumption in Law

This approach can be traced back at least to the writings of legal scholar and Supreme Court justice Oliver Wendell Holmes. In his influential 1897 essay “The Path of the Law,” Holmes argued that in trying to understand how legal rules influence human behavior, there is little point in trying to take account of either ethics or morality. Indeed, Holmes believed that “nothing but confusion of thought” could result. He urged a different approach: “if you want to know the law. . . you must look at it as a bad man, who cares only for the material consequences which such knowledge allows him to predict, and not as a good one, who finds his reason for conduct. . . in the vaguer sanctions of conscience” (Holmes 1887, 460).

Holmes went on to explain, “Take again. . . the notion of legal duty, to which I have already referred. We fill the word with all the content which we draw from morals. But what does it mean to a bad man? Mainly, and in the first place, a prophecy that if he does certain things he will be subjected to disagreeable consequences by way of imprisonment or compulsory payment of money” (459–60). Holmes’s “The Path of the Law” thus can be viewed as one of the earliest examples of the conventional law-and-economics assumption that the best way to understand how legal rules influence behavior is to assume that people treat the possibility of legal sanctions simply as a possible cost in

a rational cost-benefit analysis of whether it is personally advantageous to conform to the requirements of the law.

In adopting this approach, Holmes was embracing one of the basic precepts of nineteenth-century economic theory, that people are purely selfish actors. As Francis Edgeworth put the point, “the first principle of Economics is that every agent is actuated only by self-interest” (Edgeworth 1881, 16). John Stuart Mill in 1836 similarly viewed political economy as concerned with man “solely as a being who desires to possess wealth” (Mill 1948, 137).

## 1.2 *Homo Economicus* in Contemporary Law and Economics

In the spirit of Holmes and nineteenth-century economic thought, twentieth-century economic analysis of law continued to view Economic Man as a being focused entirely on maximizing his own material well-being. Thus it became standard practice in analyzing tort and contract law, for example, to presume that both potential injurers, and contracting parties, were selfish actors who would not hesitate to impose risks on others or breach contractual obligations whenever the anticipated personal costs were less than the anticipated personal benefits of doing so. Even in criminal law, the sanction of imprisonment was treated as a kind of “price” for criminal conduct that criminals incorporated into their personal cost-benefit analyses in deciding whether or not to commit a crime (Becker 1993). As Richard Posner summarized the standard approach in his classic treatise on the economic analysis of law, “The task of economics, so defined, is to explore the implications of assuming that man is a rational maximizer of. . . what we shall call his ‘self-interest.’ . . . The concept of man as rational maximizer of his own self-interest implies that people respond to incentives” (Posner 1998, 3–4).

## 1.3 *Homo Economicus* as a Sociopathic Actor

By assuming people always sought to rationally maximize their own wealth, conventional economic analysis was able to make human behavior seem both predictable and malleable. Pay someone to do something and he would do more of it. Impose a cost, and he would do less.

This approach gave law and economics an attractive patina of mathematical rigor. Yet at the same time, the assumption that people always pursue their own material self-interest is not only unflattering, but highly unrealistic and unempirical. Although for a variety of reasons people tend to notice selfish behavior more than they notice unselfish “prosocial” behavior (Stout 2011, 45–71), casual observation quickly reveals that people frequently violate the *homo economicus* assumption by making sacrifices to follow ethical and legal rules and to help or to avoid harming others. Thus we daily

observe large individuals waiting patiently in line rather than shoving their way to the front; we see people extend innumerable small courtesies and benefits to others; and as a general matter, we routinely observe the mass of humanity refraining from theft, fraud, and mayhem. Indeed, prosocial behavior is so common that when we find the rare individual who truly behaves like *homo economicus* and always pursues his or her own self-interest without regard for consequences to others, we label such relatively rare actors psychopaths or sociopaths (Stout 2011, 43–49).

#### 1.4 Why the Concept of Utility Doesn't Solve the Problem

One common response in the economic literature to the suggestion that *homo economicus* is a functional psychopath is to argue that economic analysis does not necessarily assume people are always selfish in the sense they maximize their own material wealth. Thus discussions of Economic Man sometimes describe him as a rational maximizer not of money or material goods, but of “utility,” a word that is broad enough to capture a wide range of desires beyond the accumulation of material wealth. (For an example, see Posner 1998, 4–5.) The next step is simply to assume that *homo economicus* sometimes gets utility from behaving ethically or from helping or avoiding harming others. This approach makes apparently unselfish behavior, like giving to charity or obeying the law when the police are not present, seem consistent with self-interest “broadly defined.”

Unfortunately, this strategy also reduces economic theory to a tautology. By defining “incentives” to include not only material goods but also such intangibles as love, esteem, and a sense of obligation, the word “incentives” has its meaning expanded to include any possible human motivation. This renders economic logic circular. The fundamental principle of economics is that people are motivated by incentives. If we then define incentives as anything that motivates people, the tautology is readily apparent (Demsetz 1999, 492; Stout 2011, 33).

This circularity is troublesome not only because it shows a lack of intellectual rigor. More worrisome, it undermines the value of economic analysis as a means of predicting behavior. As an example, consider the law of demand. The law of demand predicts that as the price of a good rises, demand for that good will decrease. Yet this prediction will only hold if we implicitly assume consumers seek to maximize their own wealth. If we were to assume instead that gas consumers want to altruistically benefit shareholders of gas companies, we would predict that a rise in gas prices would cause consumers to buy more gas rather than less. Without a strict *homo economicus* assumption, the law of demand loses its predictive power.

Thus we are left with a challenge. We know that *Homo sapiens* (as contrasted with *homo economicus*) frequently violates the prediction of purely selfish behavior by sacrificing to follow ethical rules and to help or avoid harming others. How can we incorporate this reality into economic analysis without destroying its predictive power?



## 2 EXPERIMENTAL GAMING RESULTS

---

The challenge posed above has inspired a number of scholars to adopt a behavioral economics approach in the quest to develop a model of human behavior that incorporates the empirical reality of prosociality, while still allowing economic analysis to retain most of its predictive power. They have been greatly assisted in this endeavor by the rise of the empirical methodology known as experimental gaming.

In an experimental game, human subjects in carefully controlled laboratory conditions are asked to make choices while experimenters observe how the subjects' behavior changes in response to changes in experimental conditions. One particular trio of experimental games—the *social dilemma* game, *dictator* game, and *ultimatum* game—has proven especially useful in helping behavioral economists and other experts develop a working model of when and how people will act prosocially. The literature on these three games is extensive and includes the results of hundreds if not thousands of individual treatments. From this massive data set, several consistent empirical findings have emerged (for a comprehensive review of this body of work see chapter 2 by Gächter in this volume).

### 2.1 Prosocial Behavior Is Common

The first finding is that prosocial behavior is endemic. This has been demonstrated, for example, by hundreds of social dilemma experiments. In a social dilemma, a group of players is assembled, and each player asked to choose between the “defection” strategy of maximizing his or her own wealth and the “cooperative” strategy of making a sacrifice to increase returns for the group as a whole. As in the case of the classic prisoner's dilemma, subjects in a social dilemma reap greater average payoffs if all chose to cooperate and sacrifice for the benefit of the group. However, also as in the prisoner's dilemma, the incentives in a social dilemma game are structured so that any individual player always does better by defecting, regardless of what the others in the group do (see Dawes 1980).

Hundreds of social dilemma experiments demonstrate that a large portion of players in the experiment typically will choose to cooperate and sacrifice for the benefit of the group (Stout 2011, 84–86). For example, one meta-analysis of more than 100 social dilemma experiments by sociologist David Sally found that the cooperation rate in the experiments averaged about 50% (Sally 1995, 62). Psychologist Robin Dawes and economist Richard Thaler similarly reported in a survey of single-play social dilemmas that players in the games chose to sacrifice for the benefit of the group between 40% and 60% of the time (Dawes and Thaler 1988, 187–97). Social dilemma experiments thus establish beyond reasonable dispute that people frequently act altruistically, in the sense that they elect to make a material sacrifice in order to benefit others.

## 2.2 Prosociality Includes Spite as Well as Altruism

However, altruistic self-sacrifice is not the only type of unselfish behavior people display in experimental games. This can be seen in the results of a second type of game that has been studied extensively, the ultimatum game. A typical ultimatum game has two players. One player, called the “proposer,” is given a sum of money, and then told that she can decide to give any portion that she chooses to the second player, who is called the “responder.” The responder is then presented with his own choice. He can accept the amount of money the proposer chose to give him. Or he can reject the proposer’s offer. If the responder chooses to reject the proposer’s offer, both players get nothing (see Camerer and Thaler 1995).

If *homo economicus* were asked to play an ultimatum game, the proposer would offer the minimum amount of money possible short of offering nothing at all. Then the responder would accept this token amount since rejecting it will result in her receiving nothing. Experiments consistently find that real people don’t play ultimatum games the way *homo economicus* would. Rather, in the typical ultimatum game experiment, the proposer usually offers the responder a substantial portion of the stake she has received, most frequently half. And if the proposer does not offer to share a substantial portion, the responder usually will reject the proposer’s offer (Camerer and Thaler 1995). Thus, one recent survey of ultimatum games played in a wide variety of non-Western cultures concluded that “in different countries, the majority of proposers offer 40–50% of the total sum, and about half of all responders reject offers below 30%” (Nowak, Page, and Sigmund 2000, 1773).

Ultimatum games thus demonstrate that people will not only sacrifice to benefit others; they will also sacrifice to harm them. This sort of “unselfish” behavior (meaning that it does not maximize personal wealth) is known in lay terms as vengefulness or spite. It may seem odd to describe spite as a prosocial behavior, given that its goal is to harm another. Yet as we shall see below in discussing prosocial behavior in contracting, spite can prove useful as a vehicle for motivating others to behave more prosocially.

## 2.3 People Anticipate Others’ Altruism and Spite

Finally, a third experimental game that has shed much light on the phenomenon of prosociality is the game called the dictator game. Dictator games are quite simple. Two players are involved, and the first is given a sum of money and told she can choose to distribute all of it, some portion of it, or none of it, to the second player. The second player then receives whatever the dictator is willing to share with him, no more and no less. This is of course why the first player is called the dictator (see Camerer and Thaler 1995).

A majority of subjects asked to play the role of the dictator in an anonymous dictator game will choose to share at least some portion of their initial stakes with the second player. For example, one cross-cultural study reported that dictators in different

cultures offered to donate averages of between 20% to 32% of their stakes (Henrich et al. 2004, 27). Dictator game experiments thus confirm the sort of unselfish altruistic sharing commonly seen in social dilemma games. What is interesting and significant about dictator games, however, is that the amount dictators choose to share typically is *smaller* than the amount proposers offer in ultimatum games. For example, in the cross-cultural study mentioned above, dictators' donations averaged only between 60% and 86% of the amounts that subjects in the same cultures offered when playing ultimatum games (Henrich et al. 2004). This finding suggests that the proposers in ultimatum games have an additional motive, beyond altruism, for sharing. It appears that proposers in ultimatum games anticipate that if they give a low offer, their responder will spitefully reject it.

This finding is important. It demonstrates not only that people behave altruistically and spitefully, but that people *know that other people behave both altruistically and spitefully*, and change their own behavior in anticipation. As we shall see, such second-order influences can be important in understanding how legal rules incorporate and rely on the phenomenon of prosociality.

### 3 USING EXPERIMENTAL GAMING TO MODEL PROSOCIAL BEHAVIOR

---

Taken together, the social dilemma game, ultimatum game, and dictator game offer us three basic lessons about human behavior. The first lesson is that people sometimes sacrifice to follow rules and benefit others; the second lesson is that they also sometimes sacrifice to harm others (spite); and the third lesson is that most people anticipate that others might act altruistically or spitefully.

Experimental games thus demonstrate beyond reasonable challenge that the *homo economicus* assumption of wealth-maximizing behavior can be misleading. At the same time, beyond their cautionary value, these three lessons from experimental gaming are not yet particularly useful. We want to know when people act altruistically or spitefully, and why. To appreciate the nature of the problem, consider again the 50% cooperation rate typically observed in social dilemma experiments. This finding demonstrates people often behave unselfishly. But it also demonstrates that people often behave selfishly. Otherwise we would observe cooperation rates of 100%.

Luckily, the experimental gaming literature does more than simply prove prosocial behavior occurs. It also gives us considerable insight into when, and why, people act prosocially. Researchers have been able to arrange some social dilemma experiments so that as many as 97% of subjects sacrifice and adopt a cooperative strategy that benefits their fellow players. In others, they have been able to structure the experiments so that cooperation rates drop as low as 5% (Sally 1995, 62). This is an intriguing finding, because in any social dilemma experiment, the economic payoffs are always structured

so that an individual player maximizes his or her personal wealth by selfishly defecting rather than prosocially cooperating. The economic structure of the game is fixed. Yet the behavior we observe in these games shifts, from very low cooperation rates observed in some games to very high cooperation rates observed in others.

Nor do these differing results seem to be determined by the nature of the subjects themselves. Although differences in age, gender, and cultural background have correlated with slight differences in the proclivity to adopt a cooperative strategy in experimental games, these demographic differences are typically weak (Stout 2011, 100). It appears most shifts in the behavior observed in games can be attributed to differences in the *social* conditions under which the games are played.

### 3.1 The Importance of Social Cues

In particular, three social variables seem to play consistently significant roles in determining the incidence of unselfish behavior observed in experimental games. The first social variable is instructions from authority: subjects are far more likely to cooperate in a game when the researcher instructs them to cooperate (for a meta-analysis of numerous studies finding this, see Sally 1995; for discussion see Stout 2011). Of course, we have known since the days of Stanley Milgram's infamous feigned electric shock experiments that people will harm others when instructed to do so (Milgram 1974). The interesting thing is that people will also unselfishly *help* others when instructed to do so.

A second social variable that determines whether experimental subjects behave selfishly or unselfishly is whether they believe that others in their situation would behave selfishly or unselfishly. (For an example, see Gautschi 2000.) This pattern is consistent with the well-established sociological finding that people for a variety of reasons tend to imitate others' behavior (for discussion, see Stout 2011).

Finally, a third important social variable that influences the incidence of prosociality is the magnitude of the benefits subjects believe their prosocial sacrifice will provide to others. (This is classified as a social variable because the focus is on benefits to others, rather than to oneself.) We are more likely to sacrifice to obey rules, or to help or avoid harming others, when we think the social benefits are larger rather than smaller (for a meta-analysis, see Sally 1995; for an example see Andreoni and Miller 2002).

### 3.2 The Jekyll-Hyde Syndrome

Prosocial behavior thus conforms to what could be called a Jekyll-Hyde pattern. In some social situations (buying a car, picking a stock portfolio), most individuals behave in the purely selfish fashion of Mr. Hyde, acting as if they are indifferent to consequences for others. Yet in other social situations, the same individuals act like prosocial Dr. Jekylls, sacrificing to follow rules and to help or to avoid harming others. This

duality is so deeply rooted in human nature that it can be observed with brain-imaging technology; different parts of the brain are active when people act selfishly and when they act prosocially (McCabe et al. 2001).

As indicated, among the key social variables that influence the incidence of prosocial behavior are instructions from authority, beliefs about others' behavior, and beliefs about the magnitude of the consequences for others of one's own actions. These three variables not only consistently affect the incidence of prosocial behavior in experimental games; their importance in triggering prosociality is also supported by the literature in developmental psychology, social psychology, and evolutionary theory. (For an extended discussion, see Stout 2011.) When each of these social variables is aligned to support prosociality, a very high incidence of prosocial behavior can be observed. Conversely, when the variables are aligned to support selfishness, selfishness results.

### 3.3 Some Caveats

This basic Jekyll-Hyde model is, however, subject to three important caveats that also can be seen in the experimental data. The first caveat is that it appears that a very small percentage of the population, probably only around 2% to 3%, still act like *homo economicus* in experimental games where the social variables are aligned to trigger prosociality. This finding is consistent with the observation that a similar percentage of the population can be described as psychopaths or sociopaths (American Psychiatric Association 2013). Second, the willingness to act prosocially also seems to depend to some extent on whether one perceives the persons with whom one is interacting to be members of one's "in group" whose welfare deserves some consideration. Thus the likelihood of prosocial behavior tends to decline with social distance, and may disappear entirely when dealing with those to whom one feels no social or human connection whatsoever (for an example, see Hoffman et al. 1996; for discussion see Stout 2011). The third and perhaps most important caveat is that while most people are capable of acting prosocially when social conditions support this, personal costs seem to remain relevant. The higher the personal cost of behaving altruistically or spitefully, the less likely we are to observe altruistic or spiteful behavior (for meta-analysis see Sally 1995; for discussion see Stout 2011).

## 4 SOME LEGAL IMPLICATIONS OF PROSOCIAL BEHAVIOR

---

In sum, by using the data from experimental gaming, it is possible to develop a crude model of prosocial behavior that allows us to make testable predictions about when and to what extent most people will sacrifice their own material welfare to follow rules

or to help or avoid harming other people. This allows us to create a rough-and-ready model of human behavior that is admittedly more complex than the reductionist *homo economicus* approach, but has the distinct advantage of allowing us to make behavioral predictions that are more accurate in many circumstances than the *homo economicus* approach permits. Recognizing this, a number of legal experts have begun to incorporate the empirical phenomenon of prosocial behavior into their analyses of how law influences human behavior. Some examples are offered below.

#### 4.1 Prosocial Behavior as a Supplement to or Replacement for Regulation

Perhaps the most obvious and intriguing possibility raised by the research on prosocial behavior is that it suggests that it is not always necessary, or even wise, to use fines or imprisonment (inalienability rules) to regulate human behavior and channel it in desirable directions. The right social conditions can trigger prosocial behavior even in the absence of material incentives.

One of the earliest and most influential explorations of this possibility can be found in Robert Ellickson's 1991 study of how extralegal, prosocial "norms" determined the way neighbors resolved conflicts in Shasta County, California (Ellickson 1991). Many legal scholars who have written about norms after Ellickson have implicitly chosen to retain the *homo economicus* model by emphasizing how norms are enforced by social sanctions, especially refusing to engage in transactions with norms violators ("shunning"). (For an example see Bernstein 1992.) However, Ellickson himself explicitly argued that norms were often internalized and shaped behavior through the influence of conscience rather than any consideration of material consequences. This approach not only explains why neighbors might be willing to incur a personal cost to impose social sanctions on norms violators (evolutionary biologists call such behavior altruistic punishment), but also has far more power to explain prosocial behavior among strangers in large, anonymous, urban societies. (For discussion, see Fehr et al. 2002.)

In particular, the empirical evidence on prosocial behavior suggests that even in situations where it is difficult or impossible to employ legal or social sanctions, we may be able to encourage people to act more prosocially by arranging the social variables to support prosocial behavior. That is, we can encourage people to obey the law and to help, or at least avoid harming, others, even when neither the police nor the neighbors are in sight (Dau-Schmidt 1990). A good example of this approach can be found in the work of environmental scholar Michael Vandenberg. As Vandenberg points out, individuals and households make large contributions to the problem of environmental degradation. To the extent individuals and households can be encouraged to take account of and minimize environmental damage in making consumption decisions, society can address the problem of environmental damage without having either to resort to expensive explicit regulation, or try to rely upon social interactions. Vandenberg describes the process of encouraging people to behave more prosocially

in making personal consumption decisions as environmental “norm activation” (Vandenbergh 2005). He suggests that environmental norms can be activated and individuals can be induced to take steps to avoid environmental harm by providing them with information about how their behavior damages the environment. This approach is consistent with the empirical evidence demonstrating that people are more likely to act prosocially when they understand that their behavior has significant, rather than minor, consequences for others. Vandenbergh also argues that environmental norms can be activated by providing individuals and households with information about how others are attempting to behave in a more environmentally responsible fashion. Again, this is consistent with the evidence that prosocial behavior increases when people observe others behaving prosocially.

As Vandenbergh’s work on environmental norm activation illustrates, social “framing” that manipulates social context in a fashion that encourages prosocial behavior can be employed as a valuable supplement to legal sanctions as a means of channeling behavior. Another intriguing line of the modern literature on prosociality, however looks at the possibility that when we try the opposite approach—when we add material sanctions to a context where prosocial behavior is already expected to be the norm—we sometimes see that emphasizing legal sanctions (or even rewards?) may have the unintended and undesirable consequence of diminishing the likelihood of prosocial behavior, to the point where adding legal sanction proves counterproductive. This phenomenon is called “crowding out.”

One renowned empirical demonstration of the crowding-out phenomenon was an experiment in which parents using child care centers were, for the first time, required to pay a fine when they arrived late to pick up their children. The *homo economicus* model would predict that fining late parents would reduce the incidence of lateness. Yet the experimenters observed the opposite result—after the day care centers began using fines, parents began arriving later, more often, than when there had been no fines (Gneezy and Rustichini 2000). It appears the decision to begin imposing fines on parents who arrived late changed the parents’ perception of the social context, signaling to them that the fine was the “price” for arriving late and the social situation was more like a marketplace in which they were free to act in a purely selfish fashion.

## 4.2 Prosocial Behavior and Negligence

In addition to shedding light on how we can best use regulations (inalienability rules) to change behavior, the phenomenon of prosocial behavior also carries implications for our understanding of a number of important problems in the theory of negligence. One of the most puzzling of these is the question of why we do not see runaway negligence, despite the well-recognized reality that the prospect of being held liable for personal damages, alone, is unlikely to produce optimal deterrence for rational and purely selfish actors (for discussion see Stout 2011, 157–60).



This is because a number of elements of the modern tort system work together to ensure that tort law systematically undercompensates the victims of ordinary negligence. For example, many victims choose not to sue; of those that do sue, many do not recover or recover only partial compensation because they cannot prove with sufficient certainty the identity of the tortfeasor, the fact of negligence, or the extent of their damages; in many jurisdictions, rules of contributory negligence, damages caps, or a prohibition on certain kinds of damages (e.g., pain and suffering for deceased victims) further reduce recoveries; and finally, the widespread availability of liability insurance allows many tortfeasors to avoid personal responsibility for any negligence award that might be entered against them. The inevitable result, if people were purely selfish, should be endemic negligence. Nevertheless, routine negligence does not appear to be the norm in most developed nations. This is true even in states like Canada (which has a no-fault regime for auto accidents) and New Zealand (which in 1974 abolished tort liability for personal injuries and replaced it with a state compensation scheme) (Schwarz 1994).

Making tortfeasors pay for the full costs of all losses due to their negligence is unnecessary to achieve a reasonable level of deterrence, however, to the extent societies can use social framing to encourage prosocial individuals to use care and avoid imposing unreasonable risks on others. In fact, many elements of modern societies do just that. For example, respected authorities teach that it is wrong to harm others or damage their property; we routinely observe those around us taking care to avoid harming others; and the benefits to our fellow citizens when we avoid accidentally killing or maiming them are usually obvious and highly salient. This means that even without legal sanctions for negligence, as in New Zealand, many and perhaps most individuals can still be relied upon to exercise some degree of care towards others (Stout 2011, 162–64). Tort liability can be understood as a kind of safety net or backup system to shore up prosocial behavior in situations where individuals might be tempted, perhaps because of social distance or a relatively high personal cost, to indulge in conduct that poses undue risks for others. The doctrine of negligence steps in at this point to create a marginal additional incentive, in the form of the possibility of personal liability, to tip the balance back toward prosocial behavior.

This observation sheds light on another puzzle of negligence law: why, when tort law typically undercompensates most accident victims, it deliberately overcompensates a select few who are victims of negligence so egregious the tortfeasor is found liable not just for compensatory damages, but punitive damages. Punitive damages are only awarded against the tortfeasor whom a judge or jury finds to have indulged not merely in negligence, but in “wanton negligence,” “malice,” “outrageous conduct,” or “reckless indifference” toward others. In such cases, the law gives judges and juries the discretion to require the injurer to pay damages that may be several times the magnitude of the actual harm suffered by the victim.

Some conventional law and economic scholars have criticized punitive damages on the theory that requiring injurers to pay damages in excess of the harm they cause can result in inefficient overdeterrence (Viscusi 1998). Yet both the underdeterrence typical

of normal negligence rules, and the overcompensation characteristic of punitive damages cases, make sense when we take account of the possibility that the vast majority of the population are prosocial individuals who already have a significant internal “incentive” to avoid harming others. Punitive damages, in turn, can be understood as a specialized rule that is designed to influence the behavior of those rare individuals who lack the capacity to behave prosocially and so always hew closely to the *homo economicus* model of purely selfish behavior. This explanation fits neatly with cases that have described the legal standard for punitive damages as requiring a finding that the defendant acting with “malice” or “indifference” toward others—which is of course Economic Man’s normal mental state.

### 4.3 Prosocial Behavior and Contract

As another example of how incorporating the behavioral phenomenon of prosociality into the analysis can shed light on law, consider contract law. On first inspection, the idea that prosocial behavior is important to contracting may seem odd, because contract is associated with exchange, and arm’s-length bargaining in the marketplace is one social situation in which we routinely view it as both common and acceptable for people to behave in a purely self-interested fashion. It is important to recognize, however, that contract involves *deferred* exchange, which requires one party to rely on the contractual promises of the other. As a result, prosociality that takes the form of a limited regard for the welfare of one’s contracting partner can play a significant role in making sure a contract is actually performed (O’Hara 2008; Stout 2011).

Prosocial behavior is especially important when contracts are highly incomplete. This is likely when parties are bargaining under conditions of great uncertainty, or the transaction costs associated with bargaining are high, or it is difficult for a third party like a court to verify whether or not the parties have performed their contractual promises (for discussion see Eisenberg 1995). In such cases a purely selfish and opportunistic counterparty would routinely seek to exploit the large “gaps” in the contract. Knowing this, rational parties who anticipate their counterparty would behave opportunistically would be reluctant to enter an incomplete contract with such a counterparty in the first place. The result would be to make “relational contracting” nearly impossible.

Nevertheless, relational contracting is common (Macneil 2000). This suggests that parties entering relational contracts rely not only on the contract’s terms, but also on the belief their counterparty will show some regard for their welfare beyond the minimum that can be ensured by trying to legally enforce the explicit terms of the formal contract itself. Thus an important part of the process of negotiating any relational contract will include ensuring that two preconditions are met.

First, one must select a counterparty who is capable of, and perhaps even inclined toward, prosocial behavior: no rational person would want to enter a relational contract, like a business partnership, with a sociopath. In Richard Epstein’s words, sensible

businesspeople “pick [their] partners first and worry about the contract later, not the other way around” (Epstein 1996, 11).

The second precondition is that the contracting process itself should support prosocial behavior, meaning it should present a social context that signals the parties’ relationship is cooperative and each is expected to show some regard for the other’s welfare. This second observation sheds light on the belief, common among laymen, that it is not desirable when entering an open-ended relational contract to focus too carefully on the terms of the contract. Negotiating in too much detail, and expressing too much concern for legal self-protection, can be interpreted by one’s counterparty as a signal that one does not view the relationship as one that calls for mutual good faith and trust. This shift in social context in turn influences one’s counterparty to behave less trustworthily. This may explain why, as Robert Scott has described, many complex business contracts “appear to be ‘deliberately’ incomplete” (Scott 2003). The phenomenon is similar to the reluctance of the newly engaged to negotiate prenuptial contracts.

Finally, the notion that highly incomplete contracts depend on the contracting parties’ prosocial behavior to succeed helps explain how relational contracting can work even though, when contracts are highly incomplete, courts can be unable to discern or enforce the parties’ original contracting intent. (As Eric Posner has put it, courts can be “radially incompetent” at enforcing relational contracts, Posner 2000.) Even if a party to a breached contract cannot rely on the judicial system to enforce the contract correctly, being able to drag one’s faithless counterparty into litigation, especially under the American rule where each party must pay its own attorney’s fees, is itself a form of punishment. Thus the behavioral phenomenon of spite can play a useful role in promoting prosocial behavior in relational contracting, if the parties to relational contracts anticipate that were they to indulge in opportunistic behavior, their disappointed counterparty might spitefully incur the cost of a suit merely for the personal satisfaction of knowing they have managed to impose a cost on the breaching party (Stout 2011, 196–98).

#### 4.4 Prosociality, Corporations, and Corporate Law

By examining how prosocial behavior influences regulation, negligence rules, and contracting, we have seen how prosociality is relevant to all three of the broad categories into which law-and-economics scholars divide the law (property rules, liability rules, and inalienability rules) (Calabresi and Melamed 1972). In this section we consider the particular case of an area of law that intersects with all three types of rules: corporate law.

As in the case of contract, it may seem on first inspection odd to assert that prosocial behavior is important to understanding corporate law, as nonexperts typically think of the business world as characterized by ruthless self-interest. Yet if the corporation is (as economists like to say) a “nexus of contracts,” the number of contracts involved can be enormous, and many of those contracts are highly incomplete. Thus, especially when it

comes to relationships within the firm, including the relationships between and among employees, officers, shareholders, and directors, there is every reason to suspect that when the humans who interact with and within corporations act prosocially, we may see different and often superior results than we see in corporations whose human participants always behave like *homo economicus*.

As an example, one area of corporate law in which the phenomenon of prosocial behavior has been identified by legal scholars as important is the area of fiduciary duties. (For examples see Blair and Stout 2001; Rock 1997.) In effect, the law asks fiduciaries to set aside self-interest, and behave as if their only concern was their beneficiaries' welfare. Thus directors making decisions on behalf of the corporate entity are supposed to focus only on the corporate entity's interests, and disregard their own. Of course, *homo economicus* would be incapable of such a feat. Nevertheless, that is what the law asks fiduciaries to do, apparently with at least some success.

Corporate law accomplishes this task, perhaps imperfectly, by creating a social context that explicitly calls for unselfish behavior. For example, corporate scholar Ed Rock has pointed out that even though the business judgment rule makes the fiduciary duty of care enforceable only to a very limited degree and only in extreme circumstances, courts nevertheless seek to influence the behavior of directors not by requiring them to pay damages for negligence but by "exhorting" them to behave more carefully (Rock 1997). Thus courts in general, and the Delaware Chancery in particular, seek to influence the behavior of corporate fiduciaries by acting as respected authorities giving explicit (if unenforced) instructions about what is, and what is not, appropriate conduct. Blair and Stout have argued that this approach is likely to work to at least some extent because the fiduciary duty of loyalty strictly limits directors' ability to use their corporate powers to enrich themselves. This greatly reduces the personal cost of directors' behaving prosocially: as they can't use their corporate powers to benefit themselves, they are more likely to prosocially use their powers to benefit the firm (Blair and Stout 2001).

For related reasons, the phenomenon of prosocial behavior raises questions about the wisdom of applying conventional economic analysis and the "optimal contracting" approach to executive and director compensation. Optimal contracting theory implicitly assumes both employers and employees are incapable of trusting or behaving trustworthily: each side to the employment contract is eager to opportunistically exploit the other, if given a chance. As a result, optimal contracting theory must rely on explicit ex ante incentive arrangements to control behavior. (See Bebchuk and Fried 2004, for an example of this optimal contracting approach.)

Yet the contracts between corporations and their executives and directors are almost always highly incomplete, making it difficult or impossible to eliminate all chances for opportunistic behavior (Blair and Stout 2001). For example, the business judgment rule makes it notoriously difficult to enforce the rules of fiduciary against corporate officers and directors; as another example, an empirical study of CEO employment contracts found that the majority are highly incomplete and nearly one-third of CEOs lacked any written employment contract at all (Schwab and Thomas 2006). Given this reality,

attempts at optimal contracting between corporations and their human agents may actually be counterproductive, if they “crowd out” prosociality by framing the employment relationship as one in which purely selfish behavior is both expected and accepted (Frey and Jegen 2001; Frey and Osterloh 2005). The dangers are greater still when incentive contracts create opportunities to reap large financial rewards by acting unethically or even illegally, a situation that can tempt even individuals otherwise inclined toward prosocial behavior to misbehave. Thus optimal contracting may in fact produce worse agent performance than relying on incomplete contracts premised on trust and trustworthiness (Stout 2014). In other words, in the area of director and executive compensation, adopting the conventional economic assumption of rational selfishness may not only impede the analysis, but generate counterproductive policy recommendations by making rational selfishness a self-fulfilling prophecy.

Finally, as noted earlier, social signals can be used to supplement or even replace regulation when we can use social context to encourage individuals and households to behave more prosocially. This approach may be far less successful, however, when applied to corporate rather than human entities. While corporations must act through human agents, the institution of the corporation creates its own social context that can shift the behavior of a firm’s human agents in an asocial or even antisocial direction (for an experiment confirming this possibility, see Greenfield and Kostant 2003). Thus, when we want to channel corporate behavior, it might make sense to apply different legal rules and strategies (for example, stricter regulations or treble damages rules) to corporate entities than to human entities (Stout 2011, 171–73).

## 5 CONCLUSION: THE VALUE OF INCORPORATING PROSOCIAL BEHAVIOR INTO ECONOMIC ANALYSIS OF LAW

---

For decades the *homo economicus* model of wealth-maximizing behavior has provided the explicit or implicit foundation for most economic analysis of law. Yet the empirical evidence demonstrates beyond reasonable dispute both that people often act prosocially, and that they expect prosocial behavior from others. Recognizing this empirical reality can help us better understand how and why legal rules shape behavior. This chapter has shown how taking account of prosocial behavior offers insight into a variety of legal issues touching on a wide range of property, liability, and inalienability rules.

For this reason alone, it is worth incorporating prosocial behavior into the analysis of legal and policy problems. But there may be an additional danger in relying too heavily on the *homo economicus* model. In brief, when we approach behavioral questions with the implicit assumption that people will always act selfishly, this can become a self-fulfilling prophecy. When respected authorities like economics and law experts

repeatedly assert that people are always selfish, this implies selfish behavior is always permissible and appropriate. It also suggests other people can be expected to behave selfishly. Finally, especially when the *homo economicus* model is taught in conjunction with some version of Adam Smith's famous parable of the invisible hand, it is easy to assume that because selfishness can benefit society in the market, selfishness is usually or even always socially beneficial.

The result is that overemphasizing an analytical approach that relies on the power and prevalence of selfishness has the potential to actually increase the incidence of selfish behavior. This is a troubling possibility: a number of studies have found that societies characterized by a high degree of trust and cooperation have higher levels of investment and show greater economic growth than societies in which people in fact behave more consistently like *homo economicus* (Zak and Knack 2001; Knack and Keefer 1997). To the extent that conventional economic analysis creates a social context that undermines prosocial behavior, it frustrates the ultimate goal of economic analysis: improving social welfare.

## REFERENCES

- American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed.
- Andreoni, James., and John, Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70: 737–53.
- Bebchuk, Lucian., and Jisse, Fried. 2004. *Pay without Performance: The Unfulfilled Promise of Executive Compensation*. Cambridge, MA: Harvard University Press.
- Becker, Gary S. 1993. "Nobel Lecture: The Economic Way of Looking at Behavior." *Journal of Political Economy* 101: 385–409.
- Bernstein, Lisa. 1992. "Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry." *Journal of Legal Studies* 21: 115–57.
- Blair, Margared. M., and Lyan A. Stout. 2001. "Trust, Trustworthiness and the Behavioral Foundations of Corporate Law." *University of Pennsylvania Law Review* 149: 1735–810.
- Calabresi, Guido, and Douglas Melamed. 1972. "Property Rules, Liability Rules, and Inalienability: One View of the Cathedral." *Harvard Law Review* 85: 1089–128.
- Camerer, Colin, and Richard A. Thaler. 1995. "Anomalies: Ultimatums, Dictators and Manners." *Journal of Economic Perspectives* 9, no. 2: 209–19.
- Dau-Schmidt, Kenneth G. 1990. "An Economic Analysis of the Criminal Law as a Preference-Shaping Policy." *Duke Law Journal* 1: 1–38.
- Dawes, Robyn M. 1980. "Social Dilemmas." *Annual Review of Psychology* 31: 169–93.
- Dawes, Robyn M., and Richard H. Thaler. 1988. "Anomalies: Cooperation." *Journal of Economic Perspectives* 2, no. 3: 187–97.
- Demsetz, Harold. 1999. "Rationality, Evolution, and Acquisitiveness." *Economic Inquiry* 34: 484–95.
- Edgeworth, Frances Y. 1881. *Mathematical Psychics*. London: C. Kegan Paul.
- Eisenberg, Melvin A. 1995. "The Limits of Cognition and the Limits of Contract." *Stanford Law Review* 47: 211–59.



- Ellickson, Robert C. 1991. *Order without Law: How Neighbors Settle Disputes*. Cambridge: Harvard University Press.
- Epstein, Richard A. 1996. "Contact and Trust in Corporate Law: The Case of Corporate Opportunity." *Delaware Journal of Corporate Law* 21, no. 1: 1–25.
- Fehr, Ernst, Urs Fischbacher, and Simon Gächter. 2002. "Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms." *Human Nature* 13, no. 1: 1–25.
- Frey, Bruno S., and Reto Jegen. 2001. "Motivation Crowding Theory: A Survey of Empirical Evidence." *Journal of Economic Surveys* 15, no. 5: 589–621.
- Frey, Bruno S., and Margaret Osterloh. 2005. "Yes, Managers Should Be Paid Like Bureaucrats." *Journal of Management Inquiry* 14, no. 1: 96–111.
- Gautschi, T. 2000. "History Effects in Social Dilemma Situations." *Rationality and Society* 12: 131–62.
- Gneezy, Uri, and Aldo Rustichini. 2000. "A Fine Is A Price." *Journal of Legal Studies* 29, no. 1: 1–17.
- Greenfield, Kent, and Peter C. Kostant. 2003. "An Experimental Test of Fairness under Agency and Profit-Maximization Constraints (with Notes on Implications for Corporate Governance)." *George Washington Law Review* 71: 983–1024.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis. 2004. *Foundations of Human Sociality: Ethnography and Experiments in 15 Small-Scale Societies*. New York: Oxford University Press.
- Hoffman, E., K. McCabe, and V. L. Smith. 1996. "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review* 86: 653–54.
- Holmes, Oliver Wendell. 1897. "The Path of the Law." *Harvard Law Review* 10, no. 8: 457–78.
- Knack, Stephen, and Phillip Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *Quarterly Journal of Economics* 112: 1251–88.
- Macneil, Ian R. 2000. "Relational Contract Theory: Challenges and Queries." *Northwestern University Law Review* 94: 877–907.
- McCabe, Kevin A., Daniel House, Lee Ryan, Vernon Smith, and Ted Trouard. 2001. "A Functional Imaging Study of Cooperation in Two-Person Reciprocal Exchange." *Proceedings of the National Academy of Sciences* 2001: 1662–73.
- Milgram, Stanley. 1974. *Obedience to Authority: An Experimental View*. New York: Harper & Row.
- Mill, John Stuart. 1948. "On the Definition of Political Economy." In *Essays on Some Unsettled Questions of Political Economy*, ch. 4. London: London School of Economics and Political Science.
- Nowak, Martin A., Karen M. Page, and Karl Sigmund. 2000. "Fairness versus Reason in the Ultimatum Game." *Science* 289: 1773–75.
- O'Hara, Erin Ann. 2008. "Trustworthiness and Contract." In *Moral Markets: The Critical Role of Values in the Economy*, edited by Paul J. Zak, 173–203. Princeton, NJ: Princeton University Press.
- Posner, Eric. 2000. "A Theory of Contract Law under Conditions of Radical Judicial Error." *Northwest University Law Review* 94, no. 3: 749–74.
- Posner, Richard A. 1998. *Economic Analysis of Law*. 5th ed. New York: Aspen Law & Business.
- Rock, Edward B. 1997. "Saints and Sinners: How Does Delaware Corporate Law Work?" *UCLA Law Review* 44: 1009–107.
- Sally, David. 1995. "Conversation and Cooperation in Social Dilemmas: A Meta-analysis of Experiments from 1958 to 1992." *Rationality and Society* 7, no. 1: 58–92.



- Schwarz, Gary T. 1994. "Reality in the Economic Analysis of Tort Law: Does Tort Law Really Deter?" *UCLA Law Review* 42: 377-444.
- Schwab, Stuart J., and Randall S. Thomas. 2006. "An Empirical Analysis of CEO Employment Contracts: What Do Top Executives Bargain For?" *Washington and Lee Law Review* 63: 231-70.
- Scott, Robert E. 2003. "A Theory of Self-Enforcing Indefinite Agreements." *Columbia Law Review* 103: 1641-99.
- Stout, Lynn A. 2011. *Cultivating Conscience: How Good Laws Make Good People*. Princeton, NJ: Princeton University Press.
- Stout, Lynn A. 2013. "Killing Conscience: The Unintended Behavioral Consequences of Pay for Performance." *Journal of Corporation Law* (forthcoming 2014).
- Vandenbergh, Michael P. 2005. "Order without Social Norms: How Personal Norm Activation Can Protect the Environment." *Norwestern University Law Review* 99: 1101-66.
- Viscusi, W. Kip. 1998. "Why There Is No Defense of Punitive Damages." *Georgetown Law Journal* 87: 381-95.
- Zak, Paul J., and Stephen Knack. 2001. "Trust and Growth." *Economic Journal* 111: 295-321.

## CHAPTER 9

---

# BEHAVIORAL ETHICS MEETS BEHAVIORAL LAW AND ECONOMICS

---

YUVAL FELDMAN

### 1 INTRODUCTION

---

THE last twenty years have seen a dramatic increase in the influence of psychology on the field of economics in general and on the law and economics movement in particular. As a result, significant efforts have been devoted to mapping the flaws in human cognition and examining their implications for how individuals deviate from making optimal decisions (see Jolls, Sunstein, and Thaler 1998; Korobkin and Ulen 2000; Langevoort 1998; Jolls 2007). For example, the literature has investigated how irrelevant factors of context, framing, or situation can cause individuals to make decisions that are contrary to their best interest. The literature reviewed in this chapter focuses on the importance of these automatic processes and examines how they affect people's moral and ethical behavior.

It is important to clarify at the outset the difference between behavioral law and economics (BLE) and behavioral ethics (BE). In contrast to BLE, BE has suffered from structural limitations that have reduced its ability to affect broad legal academic circles. For example, in the field of BE there is a relatively large number of founding fathers, while in BLE, there are two main founding fathers: Kahneman and Tversky. In particular, BE suffers from the simultaneous development of multiple, competing paradigms that can muddle the underlying points on which the literature agrees. These disagreements also prevent BE from being able to present consistent policy recommendations—another obstacle to its adoption within the law. Another limitation of BE lies in its reliance on mechanisms of behavioral automaticity and awareness (e.g. Bargh 1997, 2002), areas that are difficult to explore (for example, how is it possible to prove that people are unaware of their selfish intentions?), whereas classical BLE focuses on suboptimal outcomes, which are easily examined empirically. This places many of the findings of BE in an inferior methodological position relative to those of BLE.

Another body of literature that stands in contrast to BE concerns the prosocial account of human behavior (Stout 2010; Benkler 2011). According to this literature, rational choice models cannot account for our ability to cooperate and engage in prosocial behavior beyond what is in our self-interest. The present chapter takes a complementary view according to which, in some contexts, the rational choice accounts of self-interest are underinclusive when describing our self-interested behavior.

Both the BE and the literature on prosocial behavior agree on the need to take a broader view of what self-interest is, and both disagree with the notion that money is the main motivating force of people. But they do not agree on what this means: BE believes that a broad account of self-interest will reveal our tendency toward selfish action, whereas the prosocial literature argues the opposite. I do not suggest that one must look at people's selfish choices to understand their behavior. On the contrary, the arguments of BE and of this chapter offer a more complex view of what it means for a choice to be in one's self-interest and how this affects behavior. The differences are as follows: First, BE takes the broad view that many people's actions are based on self-interest in that they serve a need to maintain a positive and coherent view of a person's self. Second, BE accounts for the effect that self-interest has on our cognitive processes (for example, sight and memory), as opposed to simply looking at how self-interest affects motivation. Third, BE is more concerned with how self-interest affects us implicitly than with how it shapes our explicit choices.

Both BE and the traditional BLE literature focus on automatic processes that underlie people's decision-making. But although both literatures examine the limits of current discussion of how one's self-interest influences decision-making, BE explores the automaticity of self-interest, whereas BLE explores areas where automatic decisions undermine our self-interest (Kish-Gephart, Harrison, and Treviño 2010; Bazerman and Gino 2012, 2014).

Given the importance of intentionality to the law (for example, determining culpability in criminal law), one would expect this line of reasoning to be much more central to legal scholarship than it is. But the fact that BE has developed primarily within the field of management, not of legal scholarship, and the nascent stage of the literature have kept it from being part of mainstream legal scholarship.

Section 2 of this chapter reviews and organizes the literature into manageable themes that merit more rigorous evaluation for the purposes of legal scholarship. Section 3 examines some of the key normative issues this scholarship raises for the law. In particular, it examines the general view of BE, according to which people are often unaware of their biases and the true motivations underlying their decisions. The analysis covers both the scope of the effects of these System 1 biases and their applicability to the legal context.

It is difficult to cover all of this body of literature because it has been growing exponentially. The argument I want the reader to take away from this chapter is straightforward: if deliberative choice is limited, we must reexamine how to structure legal

interventions (e.g., incentives) in order to effectively curb unethical choices. As the critical review of the BE literature demonstrates, this task is especially difficult because of an unfortunate gap in our knowledge of automaticity and ethicality. As our review will demonstrate, we know a lot about how unethical decisions can be made without full awareness. However, we still do not know enough about how we can counteract these undesirable automatic processes.

Section 4 examines how BE should inform the law and economics scholarship. In particular, it examines the shift from addressing individuals to addressing the underlying situations, and the need to determine what serves as an incentive in a world in which people's explicit ethical choices are secondary to their implicit ethical behavior. It also examines to what extent the literature on debiasing and nudging, which are common normative reactions to the findings of BLE, is relevant to bounded ethicality. The chapter concludes with some suggestions concerning disambiguation and forced reflection, which may be able to curb the types of noncompliance uncovered by the bounded ethicality literature.

## 2 THE COMPLEX EFFECT OF SELF-INTEREST ON ETHICAL BEHAVIOR

---

### 2.1 Dual Reasoning and Self-Interest

This section describes the way BE scholars have studied the concept of self-interest. More specifically, I will analyze the different approaches to behaviors that might be influenced by self-interest in an automatic fashion

The concept of two systems of reasoning has gained popular recognition in Kahneman's book *Thinking, Fast and Slow* (2011) and it lies in the core of much of the research of BLE. The general concept differentiates between an automatic, intuitive, and mostly unconscious process (labeled System 1) and a controlled and deliberative process (labeled System 2) (see also Stanovich and West 2000; Evans 2003).

BE relies on the fact that self-interested behavior is automatic. One must carefully distinguish, however, automaticity from the related, overlapping concepts of awareness, controllability, intentionality, and attention.

Despite the fact that automatic self-interested behavior is fundamental to BE, most of the experimental evidence is inconclusive about how it operates. Even if we accept the automaticity of self-interest, we do not necessarily know the extent to which we are aware of this predisposition and whether it is malleable or even controllable. Because of the importance of self-interest for the law, it is particularly important to understand the interplay between Systems 1 and 2 in ethical decisions.

## 2.2 Current Research on Automaticity

The “automaticity of self-interest” is central to BE. Therefore, we take a step back to review the history of research on automaticity in general and on the automaticity of self-interest in particular. The researcher most identified with the effect of automaticity on behavior is John Bargh et al. (1992, 1996) and John Bargh and Chartrand (1999). In a long series of experiments, Bargh explored the meaning and significance of automatic processes. The main argument he advanced is that the majority of daily behaviors are the product of automatic and unconscious processes. He also suggested that these behaviors are below our level of awareness, although the research is inconclusive as far as this part of the argument is concerned.

Automaticity is typically examined using the technique of priming. Priming involves displaying some cue (for example, an image) to participants in such a way that they are not consciously aware of the fact that they have been exposed to the cue. In a well-known study, priming stereotypes of elderly people and of African Americans caused people to adopt slower moves and aggressive behavior, respectively. Priming techniques have also been observed in legally relevant behaviors such as competitiveness and cooperation (Bargh, Chen, and Burrows 1996; Kawakami et al. 2003; Kay et al. 2004; Doyen, Klein, Pichon, and Cleeremans 2012).

Money has also been shown to be an effective primer, both methodologically and theoretically. In a series of nine experiments, Vohs and co-authors; Vohs, Mead, and Goode (2006, 2008,) demonstrated that exposing people to subtle reminders of money completely changed their mode of behavior in the areas of cooperation and social interaction (for a related study see Kay et al. 2004). Aquino, Freeman, and Reed (2009) compared the effects of priming with such stimuli as the Ten Commandments, morality, and financial incentives on behaviors like contribution to public causes and dishonesty.<sup>1</sup> Their findings revealed an interaction between participant’s level of moral identity<sup>2</sup> and the situational prime, where people with high moral identity were more likely to change their behavior following both moral and financial primes relative to people with low moral identity. More recently, Kouchaki and coauthors (2013) demonstrated that priming with money increases the inclination to engage in unethical behavior (Yang et al. 2013). These studies show the relevance of priming to our understanding of automatic human behavior and to the law.

Another relevant example of automatic behavior is “embodiment.” In contrast to priming, which attempts to circumvent System 2 by creating a stimulus that cannot be fully processed by it, embodiment involves a physical stimulus that interacts with the individual’s body (e.g., Niedenthal et al. 2005). Kouchaki, Gino, and Jami (2014)

<sup>1</sup> Amir, Mazar, and Ariely (2008) provided an earlier demonstration of the effect of the Ten Commandments.

<sup>2</sup> Moral identity is a self-regulation mechanism that was shown to predict moral behavior. For a review of the various approaches to moral identity and to its relationships to other self-related mechanisms see Aquino and Reed (2002), Babcock and Loewenstein (1997).

showed, for example, that carrying heavy weights caused people to feel guiltier about a given behavior than people carrying lighter weights. In another example, washing one's hands was shown to reduce feelings of guilt when recalling unethical behavior (Zhong and Liljenquist 2006), and to reduce the severity of moral judgments against those who behaved unethically (Schnall, Benton, and Harvey 2008). Although these techniques are relevant to the law because they show the automaticity of some legally relevant behaviors, the most relevant question is how automatic processes affect our reasoning capabilities. The following section examines this question.

### 2.3 Motivation and Cognition

Many of the paradigms of BE are based either directly or indirectly on “motivated reasoning.”<sup>3</sup> Kunda (1987, 1990) argues that our internal motives have a significant effect on our complex reasoning skills. An example of such a skill is our ability to establish a causal connection between events. In particular, Kunda defines two distinct types of motivated reasoning, depending on whether our motive is to arrive at an accurate conclusion or at a desired conclusion.

The latter type of motivated reasoning is the key behavioral process that underlies BE. Kunda suggests that motivated reasoning affects our higher level reasoning skills, but other evidence suggests that motivated reasoning affects our memory and perception and moral reasoning skills that are highly relevant to the law (e.g. Uhlmann et al. 2009). Shu, Gino, and Bazerman (2011) showed that within very short periods of time people misremember both what they did and what they were told to do, when such misremembering allows them to believe that they had acted ethically. In particular, participants who cheated in the experiment demonstrated higher moral disengagement, and those who had read an honor code before the experiment demonstrated significantly worse recall of the code than did those who did not have an opportunity to cheat. Our compliance with the law is dependent on our remembering the relevant law; therefore the notion that memory itself could be corrupted by our self-interest becomes relevant to how we design laws.

Balcetis and Dunning (2006) used a great variety of physiological measurements, including eye tracking, to show that motivation affected visual capabilities. Their studies showed that preconscious processing of stimuli in the visual environment was affected by people's preferences, suggesting that awareness is indeed highly limited when it comes to the ability to process information that is inconsistent with one's self-interest.

<sup>3</sup> Theories such as “moral licensing” or “compensatory ethics” which focus on how people maintain their self-concept by making a change in their behavior, are an exception. Even these exceptional theories take into account some change in people's understanding of the action to maintain a positive view or receive some potential credit.

## 2.4 Awareness

The above bodies of research have implications for one of the basic questions underlying BE: even if self-interest is automatic, are we at least aware that we are behaving in a self-interested manner? The previous two subsections indicate that in all likelihood we are not. Indeed, most of the literature on BE in general, and on self-deception in particular, has found that we are unaware of the effects that System 1 has on our level of awareness to these biases. However, it seems that this discussion is far from being resolved and that there are still strong dissenting voices within this literature that support the view of conscious self-deception process (see Von Hippel and Trivers 2011).

Naturally, whether or not we are aware of these automatic processes and of the behaviors they cause has numerous legal implications. For example, it might alter our approach towards criminal responsibility. The lack of awareness should lead legal policymakers to rethink what interventions may be necessary in other domains of law as well.

## 2.5 Automaticity and Ethical Behavior

### 2.5.1 *The Automaticity of Self-Interest*

The automaticity of self-interest is one of the most basic tenets in BE, yet at the same time one of the most controversial theoretically and empirically. Moore and Loewenstein (2004) were among the first to show that self-interest and concern for others affect behavior through different cognitive systems, and that self-interest is automatic, viscerally compelling, and often unconscious. By comparing people's private beliefs and public behavior Moore et al. (2010) demonstrated that people truly believed their own biased judgments, not recognizing that their behavior was problematic.<sup>4</sup> A similar view was advanced by Gino and coauthors (2011), who demonstrated that the level of control needed to behave ethically is much higher than that following from the decision to be unethical.

A related but different mechanism of automaticity of self-interest is that of egotism. In contrast to the automaticity of self-interest, which focuses on what is good for the decision-maker, egotism focuses on the tendency of people to appreciate and unconsciously prefer things they associate with themselves. This includes various personal details such as their name and even items to which they were exposed in the past (Nuttin 1987; Jones, Pelham, Carvallo and Mirenberg 2004; Jones, Pelham, Mirenberg and Hetts 2002). The role of egotism, like that of priming and embodiment, is troubling insofar as we cannot determine how aware we are of the biasing influence of seemingly unrelated facts such as a shared birthday.

<sup>4</sup> The measurement of "private" evaluations was done by giving participants incentives to be accurate in their predictions.



### 2.5.2 *The Automaticity of Morality*

A most intense debate has been raging over whether System 1 or System 2 is responsible for making moral judgments. Jonathan Haidt (2001) and Haidt and Joseph (2004) demonstrated experimentally the role of System 1 in forming our moral reasoning. Haidt argued that moral reasoning, a product of System 2, is likely not to drive our moral judgments because the automaticity with which we evaluate targets undercuts the notion that deliberative, System 2 processes come into play in moral evaluations. He also cited behavioral evidence of the tendency to create post hoc reasons to explain events, and the importance of moral emotions in moral judgment, to demonstrate the likelihood that System 1 is responsible for moral judgments (see also Mikulincer and Shaver 2012; Haidt and Bjorklund 2007).

The views advocated by Haidt have been subject to a variety of criticisms by psychologists. For example, Cushman, Young, and Hauser (2006) argued that the intuitiveness of morality cannot be assumed across the board, (see also Pizarro and Bloom 2003) and that there is a variation in the accessibility of different moral principles in people's automatic and conscious mechanisms.

### 2.5.3 *The Debate on Whether System 1 or System 2 Is Responsible for Unethical Behavior*

In contrast to the above discussion of moral judgment, the majority view in the literature grants System 1 the leading role in unethical behavior. Epley and Caruso (2004) concluded that automatic processing leads to egocentric ethical interpretations. Similarly, Van den Bos and coauthors (2006) found support for the notion that when appraising a situation, we prefer outcomes that benefit ourselves, and only later correct to take into account fairness toward others. Using an implicit association test, Marquardt and Hoeger (2009) showed that decisions were made based on implicit rather than explicit attitudes (although they also found that implicit attitudes were correlated with choices that subjects believed to be moral). Moore and Loewenstein (2004) suggested that the effect of self-interest is automatic and associated with System 1. They wrote that "in many instances of conflict of interest, self-interest tends to operate via automatic processes whereas ethical and professional responsibilities operate via controlled processes" (p. 195).

Thus, according to the majority view, System 1 is responsible for most types of soft unethical behaviors, where, given more time and reflection, most people of standard moral character would choose to behave in an ethically appropriate manner. As a further proof of this notion, in the context of dishonesty research, Shalvi, Eldar, and Bereby-Meyer (2012) studied the time pressure effects on honesty and suggested that people's automatic reaction tends to be more dishonest (see also Gino, Ayal, and Ariely 2009; Chugh, Bazerman, and Banaji 2005).

A minority opinion in this literature holds that people's automatic responses are at least some of the time more cooperative than their deliberative responses. Greene and Paxton (2009) showed that dishonesty involves more of the control-related parts

of the brain than honesty does. In a more recent work, Rand, Greene, and Nowak (2012) showed that people's immediate response is more cooperative than their reflective response. An open question is whether it is possible to reconcile these approaches, at least partially, by suggesting that being cooperative and dishonest serves people's intuitive self-interest, especially in a social context, where appearing better seems to be the rational move from an evolutionary perspective. Although this in itself does not suggest that self-interest is deliberative, at least under some accounts of self-interest it suggests that automatic reasoning is superior to deliberative reasoning. In support of this dissenting view, Kahan (2013) demonstrated that people who score high on cognitive reflection (i.e., are more likely to use System 2 reasoning) are more likely to engage in motivated reasoning, which, as noted, is an enabling mechanism for unethical behavior.

## 2.6 We Are All Potentially Bad People

In his book about dishonesty, Dan Ariely (2012) cited an observation by locksmiths that locks are not used against "bad people," who can easily break into locked homes, but rather against good people who would nevertheless walk through an open door when they see one. The view espoused in this anecdote seems to be shared by many of the newer studies: we are all potentially bad people.

The focus on "good people" represents the growing recognition that many ethical decisions are the result of implicit, not explicit choices. Simply reviewing the titles of current papers shows how central this theme has become.<sup>5</sup> Admittedly, none of the authors listed would suggest that there are no bad people. But the contrast between saying that people are good or bad and saying that a person's individual deeds are good or bad suggests a growing recognition that many ethically relevant behaviors that were previously assumed to be choice-based, conscious, and deliberative, are anything but.

Along the same lines, Bazerman and Banaji (2004) suggested that incentives and similar concepts fail to correct a large portion of ethical behaviors because "such

<sup>5</sup> For example, "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance" (Mazar, Amir, and Ariely 2008), "Why Good People Sometimes Do Bad Things: Motivated Reasoning and Unethical Behavior" (Bersoff 1999); "How Good People Make Tough Choices: Resolving the Dilemmas of Ethical Living" (Kidder 2009); "When Good People Do Wrong: Morality, Social Identity, and Ethical Behavior" (Pillutla 2011); "Why Good People Do Bad Things: Understanding Our Darker Selves" (Hollis 2008); *Blindspot: Hidden Biases of Good People* (Banaji and Greenwald 2013). Many others do not use the term "good people" in their titles but make the same argument in the text (see, e.g., De Cremer 2011). This is also the view held by Bazerman, Loewenstein, and Moore (2002) in their studies on why "good" accountants conduct bad audits. Note that the "good people" scholarship is usually different from the type of research conducted by Zimbardo (2007) on the Lucifer effect or more recently by Valdesolo and DeSteno (2009) on character. These works generally try to explain how ordinary people end up doing evil or at least engage in gross criminal behaviors.

measures simply bypass the vast majority of unethical behaviors that occur without the conscious awareness of the actors who engage in them” (p. 111). The discussion of the lack of awareness suggests that much of the unethical behavior is done by ordinary, good people who are unaware of their unethical behavior.

Our concern, however, is not what people do but rather understanding the situational and psychological circumstances that allow them to do it. Therefore, much of the discussion in section 3 is about understanding the various mechanisms responsible for bounded ethicality.

With the focus on “good people,” an array of questions arises regarding the meaning of good and bad. Do all people have the same chance of engaging in unaware unethical behavior? Is there truly no awareness of unethicality in the case of such misconducts? And if there is no awareness, are these behaviors impossible to control or prevent? Should we replace current enforcement methods (e.g., deterrence, legitimacy) because they are ineffective when it comes to automatically triggered unethical behaviors? Current studies are far from answering these questions, and some portions of the literature contradict the notion that good people perform bad deeds.

## 2.7 Are We All Truly Good?

A central question for the law is whether people are indeed blameworthy for the types of misconducts that BE addresses. A small portion of the research in BE suggests that much of the variance in mindless unethical behaviors can be attributed to personal traits rather than simply to the power of the situation (Feldman and Smith 2013). For example, Greene and Paxton (2009) suggested that there are strong individual differences in people’s basic morality, which affect the likelihood that they will engage in automatic reasoning (note that automatic reasoning has been associated with dishonest responses).<sup>6</sup> An additional approach suggests indirectly that there are strong individual differences associated with the “moral hypocrisy,” and that people who have low self-awareness are more likely to ignore their own behavior and to judge others more harshly. None of these theories answers the underlying question of whether we are aware, and therefore responsible, for our actions, but they may call for rethinking the definition of what it means to be a good person.

<sup>6</sup> Green’s study involved an experiment under fMRI in which participants were asked to predict the outcome of a coin flip. Participants were presorted based on a questionnaire into honest, dishonest, or ambiguous, and were asked to either self-report accuracy or to write down their guesses in advance. The results showed that members of the honest group generated no more brain activity when they lost money and had no opportunity to cheat than they did when they lost money but had an opportunity to cheat. This supports the “grace” hypothesis that being honest is not the result of a “choice” involving greater brain activity.

## 2.8 Taxonomy of Main Bounded Ethicality Mechanisms

The following paragraphs review some of the main processes that underlie BE, which policymakers and academics should consider when evaluating the effectiveness of legal interventions. The first subsection reviews some of the main theoretical paradigms that account for the ways people act to protect their concept of themselves as good and coherent people. The second subsection discusses in more details the concept of moral disengagement, which is the most developed of the self-maintenance paradigms of BE.

### 2.8.1 *Self-Concept Maintenance*

Self-concept maintenance suggests that we employ motivated reasoning to bridge the dissonance between our bad deeds and the desire to view ourselves as good people. Bersoff (1999), one of the early researchers working in the area, showed in an experiment how unethical behavior is promoted when people are able to develop and maintain a biased characterization of an unethical action as being morally acceptable. The findings of Mazar, Amir, and Ariely (2008) and Ariely (2012), which indicate that people cheat only to the extent that they can maintain their self-concept of being honest, are good examples of research done in this tradition.

Dana, Weber, and Kuang (2007) have shown in a series of experiments one dominant strategy people use to maintain their self-concept while engaging in self-driven behavior—moral wiggle room. In their experiments, the dictator had to choose between a personal payoff of 6 and a personal payoff of 5. Each of these payoffs was matched with an uncertain payoff (determined exogenously by a lottery and unknown to the dictator) of 1 or 5 to the opposing player. Thus in such setting, the dictator could choose the selfish payoff (6) while convincing herself that the opposing party would receive the high payoff (5 rather than 1) as well. Before choosing between the payoffs, dictators were offered a chance to reveal (at no cost) the results of the lottery, so that they could be privy to the implications for the opposing player. Participants preferred “moral wiggle room” (that is, strategically using ambiguity to create a more favorable description of their ethical dilemma) of believing that they had not behaved unethically. In doing so, they engaged in a process of information avoidance,<sup>7</sup> leading the authors to the view that people desire to pursue self-interest while maintaining the illusion of behaving fairly (Halali, Feldman, and Banaji, in preparation).

Following a similar rationale, Shalvi and coauthors (2011) explored people’s ability to change their view of reality in order to justify their unethical behavior. They found that after merely having been shown an alternative outcome of a game, participants were willing to lie about which outcome had befallen them. Participants rolled dice under a cup so that only they could see the results. Instructed to report the results of the first roll in exchange for money corresponding to the value of the dice, participants felt

<sup>7</sup> I discuss the question of awareness in greater detail in subsection 2.4.

morally comfortable reporting the second, third, or other subsequent rolls in order to receive a higher reward.

A related theory that emphasizes the adjustment of reality to one's interests is referred to as "ethical fading" (Tenbrunsel and Messick 2004). According to this theory, by deceiving ourselves, we allow ethical concerns to fade into the background of the decision-making process, undermining their potential impact on the decision. The need to maintain our self-concept is also central to the moral hypocrisy theory. According to this theory, we can maintain our self-concept by not comparing our behavior to our preexisting moral standards. Support for this idea is found in the work of Batson and coauthors (1999), who showed that self-interested behavior decreased when participants were placed in conditions of high self-awareness. Ethical fading also suggests that people reconcile the dissonance in a different way: instead of perhaps thinking of ourselves as objectively good people, we focus on the fact that we are better than those around us.

Moral licensing theory, which suggests that we use good deeds to excuse later bad deeds, is another theory that could be seen as being based on the self-maintenance concept (see also Effron and Monin 2010). Relying on this theory, Monin and Miller (2001) found that participants who believed that they had previously established their moral credentials (in this case, lack of prejudice) felt empowered to subsequently express views that conflicted with moral norms. These findings contrast the traditional view, where those who behaved badly are more likely to do so in the future. Merritt, Effron, and Monin (2010) also showed that the reverse is true: those who have behaved well in the past are more likely to behave badly in the future if the behaviors are focused on the same dimension (see also Sachdeva, Iliev, and Medin 2009).

### 2.8.2 *Moral Disengagement*

An additional theory based on dissonance is Bandura's (1999) theory of moral disengagement. Bandura suggested that people use moral disengagement in order to justify performing inhuman actions like murder. Bandura offered eight mechanisms by which individuals can convince themselves that their actions are not immoral:

1. Moral justification used to explain why an immoral act actually has a moral purpose
2. Euphemistic labeling used to reclassify an action like "stealing" as a more innocuous act such as "shifting resources"
3. Palliative comparisons used to explain why the immoral action is a better option than its alternatives
4. Displacement of responsibility accomplished by blaming a superior who is believed to have ordered the immoral act
5. Diffusion of responsibility by stating that the immoral individuals were merely part of a group in which everyone was performing the immoral action, or that their part of the action was not immoral
6. Trivializing the consequences of the act to make it seem more acceptable

7. Dehumanizing the victim to render the action acceptable
8. Blaming the victims to render the immoral action acceptable by claiming that they provoked it

In a follow up study, Moore and coauthors (2012) applied the eight categories to explain self-deception in the context of employment. Broadly, Moore and her coauthors proposed that moral disengagement has three important effects on institutional corruption. First, it can make unethical decisions easier by reducing psychological discomfort in making such decisions. Second, because moral disengagement excludes moral values from consideration, it expedites unethical behavior by freeing up cognitive resources to work on the unethical goal. Finally, since these actions promote the interests of the corporation, individuals who morally disengage rise in the corporation, perpetuating unethical behavior.

Another taxonomy of all available mechanisms of self-deception was proposed by Ayal and Gino (2011). Ayal and Gino analyzed much of the existing literature and suggested a list of mechanisms that allow people to behave dishonestly. Among other things, they discuss the concept of moral cleansing, which they describe as an attempt to rid oneself of negative feelings after committing an unethical act by mental appeasement. This controlled process enables individuals to distance themselves from transgressions by engaging in either overt practices, such as washing hands and going to confession, or more covert ones, like moral licensing.

## 2.9 Social Norms and Fairness

Finally, there are two bodies of literature that might be seen as having the potential to limit the “damage” to society from mindless unethicality—social norms and fairness. Indeed, these two concepts are highly studied in behavioral law and economics context, as curbing some of the negative effects of reliance on self-interest (for a review and discussion of the limitations of these notions, see Feldman and Tyler 2012). In the context of BLE, the curbing of the effect of self-interest on behavior is related to limiting its conscious effects on behavior. For example, in a seminal study, Kahneman, Knetsch, and Thaler (1986) put forward the idea that people’s sense of fairness could overcome their desire to act selfishly. Using survey data on the different roles people play, they showed the importance of maintaining a sense of fairness. The BE literature, however, has responded with a relatively intuitive counterargument: fairness is a concept that is highly susceptible to self-interested interpretations, casting doubt on its ability to counter self-interest effects. For example, Thompson and Loewenstein (1992) have shown that people were more likely to remember information that was related to their own position, leading to the fact that their view of what is fair was aligned with their interest in how a settlement should look like. Babcock and Loewenstein (1997) showed a similar effect when people predicted the judicial decision in their case, thus reducing the likelihood of an out-of-court settlement.

Similarly to fairness, social norms can affect the likelihood that people will engage in unethical behavior. The effects of social norms have been incorporated into law and economics based on rational choice approaches, taking into account aspects such as reputation, expressive effects, shaming, and social sanctioning. Cooter (1998), for example, developed an economic theory of how the expressive values of law can shape social norms and individual preferences. Ellickson (1998) described the embrace by law and economics of status and social norms as mechanisms for the informal enforcement of the law. In another example, looking at trade secret law, Feldman (2009) found experimental support for the notion that the expressive function of law operates through a combination of intrinsic and extrinsic forces.

Shaming in particular received much attention in the law and economics literature. Kahan and Posner (1999) examined how shaming penalties can work to shape preferences against crime and express social norms. While indeed some BLE scholars have recognized the process of internalization and preference change, rather than merely accounting for costs and benefits (see also chapter 10 by Bilz and Nadler in this volume), the BE literature took the further step of actually exploring what factors make internalization more feasible. Scholars in the BE literature attempt to understand the factors that will enhance the automatic internalization of social norms. For example, Gino, Ayal, and Ariely (2009) showed that the effect of unethical group norms on people's inclination to engage in dishonest behavior strongly depends on the salience of group identity. In a more thorough examination of this psychological mechanism, Gino and Galinsky (2012) studied the effect of psychological closeness on the likelihood that social norms cause people to engage in unethical behavior. For instance, they showed that the likelihood that an unethical norm will lead to a change in one's ethical decision-making is highly depended on the level of psychological closeness of the participant to the unethical individual.

Another aspect of social norms with regard to which the BE literature innovates over earlier accounts is the asymmetric influence of unethical and ethical norms on behavior. Some research suggests that given the tendency for self-concept maintenance, people are more attentive to unethical behavior of others in order to subsequently justify their ethical missteps (Barkan et al. 2012; Epley and Dunning 2000; Cooter, Feldman, and Feldman 2008).

## 2.10 Concluding Remarks

The common theme in the literature on BE is that unethical behaviors are not the product of explicit choices to do wrong but rather are largely the product of System 1, mindless choices. A closer look at some of the mechanisms described in the bounded ethicality literature, however, shows that in many cases there is awareness of the misconduct, and yet we still regard the behavior as "bounded" because of mechanisms such as rationalization, or lack of attention, which prevent individuals from acting ethically. Another common theme in the literature is the need to protect our self-image,



to resolve the dissonance between believing that we are good people and our desire to maximize self-interest. Resolving this dissonance is reflected in many of the studies reviewed in this section.

Although most scholars believe that self-interest is automatic, the lack of conclusive research about whether we can control behavior despite its automaticity, or at least be aware of the behavior and compensate for it by engaging in behaviors that would curb its negative effects, casts doubt on whether those who engage in these automatic, unethical behaviors are still good people.

Above we have reviewed a body of literature with great potential for shaping legal policymaking. At the same time, we recognize a lack of consensus on many aspects of bounded ethicality, including competing and redundant underlying explanations of the relevant behaviors. Furthermore, the literature contains several theoretical and methodological limitations, including abstract experimental designs that do not necessarily mimic real-life dilemmas, and limited accounting for individual differences. There is also limited research about the long-term effects of many of the celebrated BE experimental manipulations. As I have shown, the methodological difficulty of BE derives from the fact that awareness, notoriously difficult to measure, is a crucial aspect in this field: awareness may dictate whether individuals engage in self-interested behaviors intentionally or unintentionally.

The research described above renders the normative part of this chapter especially challenging. Some theories, such as those related to fairness and social norms, suggest possible solutions, but the questions raised by the literature are far more numerous than the answers currently offered. Naturally, the stronger the case that unethical acts are committed by people who do not realize they are behaving unethically, the more limited the law's ability to change their behavior. The next section discusses these challenges.

### 3 THE NORMATIVE IMPLICATIONS OF BEHAVIORAL ETHICS

---

It seems to be a straightforward argument that individuals' limited awareness of their bad behavior requires a substantial revision in the design of legal rules that govern them (Feldman and Smith 2014). Indeed, some of the leading scholars in BE were quick to point out that classical intervention techniques, such as penalties, are rendered ineffective if ethical decisions are produced without awareness by System 1 (Banaji, Bazerman, and Chugh 2003). While the management literature is burgeoning with studies offering various prescriptions for organizations to respond to this concern, the legal scholarship has mostly ignored this need. As in the first parts of the chapter, I use "traditional" BLE as a baseline for evaluating the ways in which policymakers may be able to account for people's automatic reasoning either directly (e.g., the nudge

approach) or indirectly, by forcing people to overcome their automatic reasoning by means of System 2 (see also Amir and Lobel 2012).

Overall, according to the normative arguments presented in this part of the chapter, the classical dichotomy of intrinsic versus extrinsic motivation in law may have to be replaced by a different, broader dichotomy that combines the controlled and automatic modes of cognition underlying ethical reasoning and motivation. In the next part, I discuss the areas of law in which BE could be relevant, and attempt to account for the dichotomy between the common bad deeds of good and bad people and the way the law should respond to both.

### 3.1 The Nudge Approach

Following Sunstein and Thaler's influential book on the use of nudges to cope with bounded rationality, some scholars attempted to use the nudge approach in the context of BE as well. For example, Shu, Gino, and Bazerman (2011) showed that people were more honest and less likely to engage in moral forgetting and moral disengagement if they signed an agreement at the top of the document, making salient the need to be honest in completing the document, than if they signed it at the end of the document, at which point the decision on whether to be honest or lie has already been made. The rationale behind these set of findings is that the act of signing in itself triggers morally desirable behavior and when the signature happens only at the end, it is too late to change people's behavior in filling the document. While the strong effects demonstrated in this study might be attributed to the fact that the "sign first, write later" was seen by participants as an unusual practice, rather than simply because of the signature, it suggests the potential for nudge-related innovation in legal policymaking.

Another nudge approach is illustrated by Bohnet, Bazerman, and Van Geen (2012), who showed that when people evaluate others in a between-subjects comparison rather than in a within-subject comparison, negative stereotypes regarding race became less relevant. The authors believe that comparing multiple candidates requires more deliberative, System 2 reasoning than making a simple yes-or-no evaluation of a single candidate, which is decided by System 1 and is therefore more prone to be biased. Thus, when people need to decide between two candidates at the same time, their System 2 is more likely to monitor and reduce the potential disruptive effect of stereotypes of their decision-making.

Amir and Lobel (2008) examined the limitations of using nudge-like approaches in trying to solve policy problems. In their review of Ariely's and Sunstein's books on irrationality and nudge, respectively, they noted that if policymakers are to implement successful policy solutions, they need to do a more thorough job in classifying the mechanisms underlying behavioral biases. For example, they should properly determine whether bounded ethicality or bounded rationality is at issue. If bounded rationality is the problem, the focus should be on simplifying their choice. If, on the other hand, bounded ethicality is at issue, their focus should be on eliminating justifications for inappropriate behavior.

Although some of the normative suggestions outlined in this chapter could be seen as related to the “nudge” approach, there is a limit to the ability of these tactics to lead to a sustainable change. In the original nudge context of behavioral economics, people prefer to save money, quit smoking, choose the best schools for their children, enroll in health programs, lose weight, and avoid taking too expensive loans. In other words, the nudge, at least in those more paternalistic contexts,<sup>8</sup> works in part because it helps people achieve the goals they already have formulated.

As noted earlier, in an ethical context the situation is very different: to a large extent, especially when enforcement is limited, people benefit from cutting corners, avoiding taxes, and underperforming in a contract, but they delude themselves into believing that they have not misbehaved. Thus, if the objective is to drive them away from their self-interest, we are likely to encounter significant resistance to nudge-like approaches. We must note an additional concern regarding the face validity of some of the nudge approaches: often the studies examine ethical shifts produced by atypical cues. For example, Gino and Desai (2012) found that cues from childhood, such as soft toys and nursery rhymes, decrease unethical behavior in participants. Although studies such as these contribute greatly to understanding how people make ethical decisions, there are obvious limitations to their applicability in the legal context. Indeed, given that this research is led by management scholars, some face validity issues are less of a concern because organizations are more likely to be able to engage in soft nudge-based enforcement of unethical behaviors than government bodies, which need to maintain its legitimacy. Thus, for example, while using teddy bears or priming the Ten Commandments to enhance honest behavior might be seen as a clever ethical nudge by an employers, it is unlikely to be seen as such when done by the IRS.

### 3.2 Debiasing

Debiasing refers to attempts aimed at helping people overcome their biases by making them reevaluate their errors. For a theoretical and applied analysis of debiasing through the law, in such fields as litigation, employment, and consumer protection, see Jolls and Sunstein (1996) and chapter 28 by Sunstein in this volume. In the same vein, some scholars began to examine the possibility of debiasing people’s ethical decisions. The most elaborate examination was offered by Tenbrunsel and coauthors (2007), dealing with “ethical mirage.” The authors focus on the biases responsible for the awareness gaps between the ethicality of what people do, what they want to do, and what they think they do. In particular, they discuss the concept of a “should self,” which is capable of describing level-headedly how we ought to act and try to understand why it does not play greater role in actual decisions. The authors suggest various cognitive practices

<sup>8</sup> For examples of nudges that are used to improve the general welfare, rather than one’s own benefit see Kuenzler and Kysar, this volume.

that will allow better communications between the various parts of the self. These include detailed planning of the situation, practicing how to deal for potential sources of biases and active listening to the “should” self during the decision-making process. Similarly, Schurr and coauthors (2012) suggested broadening individuals’ decisions and forcing them to plan future choices, making it difficult for them to engage in self-deception to justify decisions based on self-interest. The concern in all of these suggestions is that people themselves have little incentive to adopt these solutions because they run counter to their self-interest (Schurr et al. 2012). Hence, for these practices to curb mindless unethical behavior, they need to be administered by organizations and states. This situation is different from, for example, classical BLE implications, where litigants are being exposed to their true winning chances in courts. In such situation the process of the debiasing could save them money and lead them to prefer settlements to overly optimistic and hence costly litigation.

## 4 TENTATIVE REFLECTION ON POLICYMAKING

---

The previous part discussed the broad ideas of nudges and debiasing as possible normative reactions to the challenges of BE. Below I explore some concrete research directions for legal policymaking that are sensitive to the unethical behavior associated with automatic processes.

### 4.1 Statistical Unethicality

Given the difficulty of determining individuals’ awareness of the unethical nature of their decisions, BE suggests focusing on aggregating people’s decisions as proof of wrongdoing. This view can borrow by analogy from the work in the area of employment discrimination, where the inability to penetrate people’s mind (Krieger 1995) has led in some cases to an aggregated approach to evaluating decision-making. Beginning with *Griggs v. Duke Power Co.*, (1971) the US Supreme Court has recognized that although it is not mandated that the workforce of a company should replicate the composition of the general population, statistical disparity between the two can be used as compelling evidence of employment discrimination under a disparate impact theory. According to this theory, even if it is impossible to prove that the employer intended to treat candidates differently, the fact that the employer used criteria that resulted in discrimination against a class of individuals is sufficient to establish an illegitimate discrimination (Shoben 1983).

Similarly, in the realm of bounded ethicality, rather than assuming bad intentions where we cannot prove them, it may be possible to collect data on ethical decisions over

time and create criteria to be applied if the aggregation of behaviors indicates that one should have been aware of the negative effect of her actions. For example, misuse of office supplies or the improper acceptance of gifts may be considered misbehavior even if any individual instance of such conduct is merely questionable. A sufficient number of marginal instances can warrant sanctioning regardless of the actor's intent. Important jurisprudential work remains to be conducted to justify increasing one's responsibility for one event based merely on the fact that it has been repeated. However, given the difficulty of determining responsibility for isolated events, and the ability of System 2 to predict the likelihood that such unethical act reoccurs, a solution of this type may be necessary and appropriate.

## 4.2 Disambiguation

Many of the theories of bounded ethicality reviewed above suggest that legal ambiguity causes people to engage in wrongdoing (see also Feldman and Harel 2008). Under the assumptions of rational choice, ambiguity and uncertainty are expected to increase compliance, as people tend to be risk averse (Garoupa 2003; Logue 2006). Behavioral ethics suggests otherwise.

One of the principal techniques that people use to self-justify unethical behavior is to employ the mechanisms listed in section 2.8 above in order to construct self-serving interpretations of legal and organizational requirements. For example, Feldman and Teichman (2009, 2011) showed that people may use legal ambiguity strategically to generate a self-serving interpretation of what is required from them by laws or contracts. Feldman, Halali, and Banaji (in preparation) found these types of interpretations of legal imperatives to be sensitive to the individual's type of interest (monetary vs. status), level of awareness, and amount of cognitive resources available to the individual (following manipulation of ego depletion and time pressure). As expected, people's need to feel that they are following the letter of the law on one hand, and their motivation to enhance their self-interest on the other, give rise to the concern that they would routinely engage in self-serving interpretation of the legal norms.

The need to avoid ambiguity may be related to other areas of law, where the lesson of BE is that at times reduced fear of behaving unethically produces more rather than less unethical behavior. For example, in the context of conflicts of interest, according to rational choice predictions we should be most concerned in situations that present the greatest conflict. But BE suggests that the opposite may be true: if there is ambiguity about the existence of a conflict, individuals will be tempted to deceive themselves that no such conflict exists (Moore, Tetlock, Tanlu, and Bazerman 2006). In clear-cut situations individuals do not have such an excuse, and therefore they are less likely to misbehave. Thus, the law should seek to avoid the creation of ambiguous partial solutions that solve only part of a problem (e.g., disclose conflict of interest but do not eliminate it (Cain, Loewenstein, and Moore 2005), for fear of creating exploitable situations.

Along the same lines, although much of the research on conflicts of interest and corruption tends to suggest that money is more easily transferable and therefore more dangerous to the integrity of society than other sources of conflict, such as prestige and loyalty, some of the research on self-deception suggests that the opposite may well be true. Most people recognize that taking money from various entities limits their ability to remain objective. In contrast, the influence of nonmonetary interests such as prestige, esteem, loyalty, and competence is less likely to trigger an alert by System 2, and therefore these interests are more likely to have an automatic effect on people, unconstrained by System 2 (Babcock and Loewenstein 1997).

### 4.3 Reflection and Accountability

Bazerman and Terbenus's (2011) examination of blind spots has called attention to the limited-awareness processes in which people engage. The classic response to the limited awareness problem is to force people to reflect on what they do, making sure that they cannot brush aside the unethicity of their behavior using any of the numerous rationales reviewed in section 2 above. Indeed, research on the interplay between System 1 and System 2 suggests that accountability may prevent people from overrelying on their automatic reasoning (Tetlock and Lerner 1999). Unfortunately, in many legal contexts in which some reporting is necessary the opposite is true: forms are constructed to promote quick, thoughtless binary decisions, leaving little room for open-ended reflection and thought (e.g., taxation, disclosures made when selling a car or a house, customs declarations, financial reports by businesses, reports by oil and gas companies of payments to foreign countries related to processing minerals). In many cases, all matters that individuals or corporations may decide not to report will never be detected, and if they are revealed they are usually regarded as honest mistakes (compare with Kahan 1997). Creating situations in which people's dilemmas should be reported explicitly in these forms may make it possible for people to reflect about what they need to do. For example, it may be possible to include open-ended rubrics on tax forms, where taxpayers would be required to indicate the areas about which they feel uncertain.

#### 4.3.1 *Regulating Situations Rather Than Incentives*

Behavioral ethics also highlights the need for changing the underlying situations individuals face, which can shape both explicit and implicit choice, as opposed to changing the pertinent incentives, which can only shape their conscious decisions (Aquino et al. 2009). Rather than changing the social cues, as nudge-like approaches do, it is necessary to shape and curtail the biases that determine the way in which people approach dilemmas even before consciously exploring how to solve them. The recommended approach is to force people to recognize when they engage in bad behavior. If people often engage in unethical behavior because they are unaware of its nature, making them recognize their unethicity should help curtail this type of behavior (Haisley and Weber 2010; Hsee 1995).

Trevino (1986), one of the pioneers of business ethics, focused on the importance of situations and business environments in facilitating unethical behaviors, above and beyond any characteristics of the individuals themselves. This focus has been further emphasized in the new BE literature. In a comprehensive review, Bazerman and Gino (2012) outlined various characteristics of situations that are likely to increase unethical behavior, such as the extent to which the situation allows for the rationalization of unethical behaviors. Thus, for example, they outline situational factors such as role-playing (e.g. the Stanford Prison Experiment), visual factors such as the presence of graffiti on walls, and social factors like relationship with a wrongdoer. People also appear to be more likely to behave dishonestly when they face losses (Kern and Chugh 2009), especially if behaving unethically can help them turn the loss into a gain (Shalvi, Eldar, and Bereby-Meyer 2012). In a world that pays attention only to deliberate unethicality, law and economics models based on incentives are more than enough to change behavior in ethical contexts. But the more we understand that people's conscious choices are responsible for only part of their unethicality, the more regulatory resources we must allocate to creating situations that reduce all forms of unethicality and legal incompliance.

#### 4.3.2 *Blind Spots and Ethical Training*

Bazerman and Tenbrunsel (2011) suggested a comprehensive set of solutions for policy-makers based on insights derived from the BE literature. The authors suggested various techniques for people to become aware of their limited ability to recognize their ethical failings, and ways of remedying this situation. For example, focusing on the aforementioned concepts of the "should" rather than the "want" self, Bazerman and Gino suggested that by considering beforehand what "want" desires may come into play at moments of decision-making, individuals can better prepare to resist these desires and instead implement a decision based on their ethically sound "should" preferences. The authors use the example of a common interview question: What pay does a competing employer offer? The "want" self is likely to wish to inflate the number, encouraging the applicant to lie to the potential employer. By anticipating this, one can come up with a more acceptable answer, such as "I'm not comfortable sharing that information," which serves self-interest but also does not violate moral rules.

Similarly, in a series of experiments, Devine and coauthors (2012) showed that it is possible to train people to be less affected by their implicit racial biases. At a minimum, these experiments support the claim that even if we are unaware of our self-interested behaviors, we may be able to control them.

## 5 CONCLUSION

---

Behavioral ethics makes an important contribution to the BLE literature. For the most part, this literature supports the view that the self-serving effects of motivation on cognition allow people to do harm when it serves their self-interest and at the same



time not feel guilty about it. Whereas according to the traditional behavioral literature biases prevent people from realizing their self-interest, the BE literature highlights countervailing biases that prevent people from understanding that their behaviors are self-interested and unethical.

Uncovering these biases is especially important because society is being harmed by nondeliberate bad deeds, while state interventions, as well as interventions called for by the law and economics literature, do not target these implicit behaviors. At the same time, as the present review suggests, the ability of the current literature to make concrete suggestions for the law is limited. Many important aspects are still being debated in the literature, both theoretically and methodologically. Within the concept of bounded ethicality, the interrelations between automaticity, awareness, and controllability are still the subject of controversy, and potential solutions are elusive. Furthermore, we know more about the effect of System 1 on System 2—which is of descriptive interest—than about the effect of System 2 on System 1—which is of greater normative interest. Even one of the most momentous questions for the law—are we intuitively good or bad—seems to be affected by context more than was previously assumed, although many scholars think otherwise.

Finally, the current literature raises various methodological issues. For example, most of the findings in BE are collected from laboratory experiments, which accounts for short-term effects of various ethical manipulations. For the most part, the law is more interested in the long-term effects of these practices. Most of the research in BE has been in the area of micromanagement, which is not entirely suitable for incorporation into legal theory and policymaking. Although preliminary abstract and concrete arguments are offered for the consideration of legal scholars and policymakers, at the conclusion of this chapter one can only hope that further research conducted by legal scholars will enrich the much-needed behavioral law and economics literature and make it more inclusive.

## ACKNOWLEDGMENTS

---

I wish to thank Amos Schurr, Shaul Shalvi, and the participants of the Law and Market Behavior Conference at Notre Dame University, April 2013, for their comments. I am also thankful to Sarit Hautka, Einav Tamir, Troy Schuler, and Peter Grogan for research assistance. Special thanks to Doron Teichman and Eyal Zamir for their detailed and insightful editorial comments.

## REFERENCES

---

- Amir, O., and O. Lobel. 2008. Stumble, predict, nudge: How behavioral economics informs law and policy. *Columbia Law Review* 108(8):2098–137.

- Aquino, K., D. Freeman, and A. Reed III. 2009. Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology* 97:1–123.
- Aquino, K., and A. Reed III. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology* 83:1423–40.
- Ariely, D. 2012. *The (Honest) Truth about Dishonesty: How We Lie to Everyone—Especially Ourselves*. New York: HarperCollins.
- Ayal, S., and F. Gino. 2011. Honest rationales for dishonest behavior. *The Social Psychology of Morality: Exploring the Causes of Good and Evil*. Washington, DC: American Psychological Association.
- Babcock, L., and G. Loewenstein. 1997. Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives* 11(1):109–26.
- Babcock, L., G. Loewenstein, S. Issacharoff, and C. Camerer. 1995. Biased judgments of fairness in bargaining. *American Economic Review* 85(5):1337–43.
- Balcetis, E., and D. Dunning. 2006. See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology* 91(4):612–25.
- Banaji, M. R., M. H. Bazerman, and D. Chugh. 2003. How (un) ethical are you? *Harvard Business Review* 81(12):56–65.
- Banaji, M. R., and A. G. Greenwald. 2013. *Blindspot: Hidden Biases of Good People*. New York: Delacorte Press.
- Bandura, A. 1999. Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review* 3(3):193–209.
- Bargh, J. A. 1997. *The Automaticity of Everyday Life*. Mahway, NJ: Lawrence Erlbaum.
- Bargh, J. A. 2012. Priming effects replicate just fine, thanks. *Psychology Today blog*. May 11. Available at <http://www.psychologytoday.com/blog/the-naturalunconscious/201205/priming-effects-replicate-just-fine-thanks>.
- Bargh, J. A., S. Chaiken, R. Govender, and F. Pratto. 1992. The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology* 62(6):893–912.
- Bargh, J. A., and T. L. Chartrand. 1999. The unbearable automaticity of being. *American Psychologist* 54(7):462–79.
- Bargh, J. A., M. Chen, and L. Burrows. 1996. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology* 71(2):230–44.
- Bargh, J. A., and M. J. Ferguson. 2000. Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin* 126(6): 925.
- Barkan, R., S. Ayal, F. Gino, and D. Ariely. 2012. The pot calling the kettle black: Distancing response to ethical dissonance. *Journal of Experimental Psychology* 141(4):757–73.
- Batson, C. D., E. R. Thompson, G. Seuferling, H. Whitney, and J. A. Strongman. 1999. Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality And Social Psychology* 77(3):525–37.
- Bazerman, M. H., and M. R. Banaji. 2004. The social psychology of ordinary ethical failures. *Social Justice Research* 17(2):111–15.
- Bazerman, M. H., and F. Gino. 2012. Behavioral ethics: Toward a deeper understanding of moral judgment and dishonesty. *Annual Review of Law and Social Science* 8:85–104.
- Bazerman, M. H., G. Loewenstein, and D. A. Moore. 2002. Why good accountants do bad audits. *Harvard Business Review* 80(11):96–103.

- Bazerman, M. H., and A. E. Tenbrunsel. 2011. *Blind spots: Why We Fail to Do What's Right and What to Do about It*. Princeton, NJ: Princeton University Press.
- Benkler, Yochai. 2011. *The Penguin and the Leviathan: How Cooperation Triumphs over Self-Interest*. New York: Crown Business.
- Bersoff, D. M. 1999. Why good people sometimes do bad things: Motivated reasoning and unethical behavior. *Personality and Social Psychology Bulletin* 25(1):28–39.
- Bohnet, I., M. Bazerman, and A. Van Geen. 2012. When performance trumps gender bias: Joint versus separate evaluation. Harvard Business School Working Paper no. 12-083. Available at <http://www.hbs.edu/faculty/Publication%20Files/12-083.pdf>.
- Cain, D. M., G. Loewenstein, and D. A. Moore. 2005. The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *Journal of Legal Studies* 34(1):1–25.
- Camerer, C., G. Loewenstein, and D. Prelec. 2005. Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature* 43(1):9–64.
- Chugh, D., M. H. Bazerman, and M. R. Banaji. 2005. Bounded ethicality as a psychological barrier to recognizing conflicts of interest. Pp. 74–95 in *Conflict of interest*, edited by D. A. Moore, D. M. Cain, G. Loewenstein, and M. H. Bazerman. Cambridge: Cambridge University Press.
- Cooter, R. D. 1998. Expressive Law and Economics.
- Cooter, R. D., M. Feldman, and Y. Feldman. 2008. The misperception of norms: The psychology of bias and the economics of Equilibrium. *Review of Law and Economics* 4(3):Article 7.
- Cushman, F., L. Young, and M. Hauser. 2006. The role of conscious reasoning and intuition in moral judgment testing three principles of harm. *Psychological science* 17(12):1082–89.
- Dana, J., R. A. Weber, and J. X. Kuang. 2007. Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory* 33(1):67–80.
- Desmet, P. T. M., D. De Cremer, and E. van Dijk. 2011. In Money we Trust? The Use of Financial Compensations to Repair Trust in the Aftermath of Distributive Harm. *Organizational Behavior and Human Decision Processes* 114(2):75–86.
- Devine, P. G., P. S. Forscher, A. J. Austin, and W. T. L. Cox. 2012. Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology* 48(6):1267–78.
- Doyen, S., O. Klein, C. L. Pichon, and A. Cleeremans. 2012. Behavioral priming: It's all in the mind, but whose mind? *PLoS One* 7(1): e29081.
- Effron, D. A., and B. Monin. 2010. Letting people off the hook: When do good deeds excuse transgressions? *Personality and Social Psychology Bulletin* 36(12):1618–34.
- E. Halali, Y. Feldman, and M. Banaji. (In preparation). Legal ambiguity and self interest: The role of dual-process interpretation in implicit corruption.
- Ellickson, R. C. 1998. Law and economics discovers social norms. *Journal of Legal Studies* 27:537–52.
- Epley, N., and E. M. Caruso. 2004. Egocentric ethics. *Social Justice Research* 17(2):171–87.
- Epley, N., and D. Dunning. 2000. Feeling “holier than thou”: Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology* 79(6):861–75.
- Evans, J. St. B. T. 2003. In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences* 7(10):454–59.
- Feldman, Y. 2009. The expressive function of the trade secret law: Legality, cost, intrinsic motivation and consensus. *Journal of Empirical Legal Studies* 6(1):177–212.

- Feldman, Y., and A. Harel. 2008. Social norms, self-interest and ambiguity: An experimental analysis of the rule vs. standard dilemma. *Review of Law & Economics* 4(1):81–126.
- Feldman Y., and Smith, H. 2014. Behavioral Equity. *Journal of Institutional and Theoretical Economics*, 170(1) pp. 137–159.
- Feldman, Y., and D. Teichman. 2009. Are all legal probabilities created equal? *NYU Law Review* 84(4):980–1022.
- Feldman, Y., and T. R. Tyler. 2012. Mandated justice: The potential promise and possible pitfalls of mandating procedural justice. *Regulation and Governance* 6(1): 46–65.
- Feldman, Y., and D. Teichman. 2011. Are All Contractual Obligations Created Equal? *Georgetown L.J.* 100(1):5–51.
- Fine, C. 2006. Is the emotional dog wagging its rational tail, or chasing it? *Philosophical Explorations* 9(1):83–98.
- Garoupa, N. 2003. Behavioral economic analysis of crime: A critical review. *European Journal of Law and Economics* 15(1):5–15.
- Gino, F., S. Ayal, and D. Ariely. 2009. Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science* 20(3):393–98.
- Gino, F., and S. D. Desai. 2012. Memory lane and morality: How childhood memories promote prosocial behavior. *Journal of Personality and Social Psychology* 102(4):743–58.
- Gino, F., and A. D. Galinsky. 2012. Vicarious dishonesty: When psychological closeness creates distance from one’s moral compass. *Organizational Behavior and Human Decision Processes* 119:15–26.
- Gino, F., and J. Margolis. 2011. Bringing ethics into focus: How regulatory focus and risk preferences influence (un)ethical behavior. *Organizational Behavior and Human Decision Processes* 115(2):145–56.
- Gino, F., D. Moore, and M. Bazerman. 2009. No harm, no foul: The outcome bias in ethical judgments. Harvard Business School Working Paper no. 08-080. Available at <http://www.hbs.edu/faculty/Publication%20Files/08-080.pdf>.
- Gino, F., M. Schweitzer, N. Mead, and D. Ariely. 2011. Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior and Human Decision Processes* 115(2):191–203.
- Greene, J. D., and J. M. Paxton. 2009. Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences* 106(30): 12506–11.
- Griggs v. Duke Power Co.* 401 U.S. 424 (1971).
- Guthrie, C. 2003. Prospect theory, risk preference, and the law. In “Symposium: Empirical Legal Realism: A New Social Scientific Assessment of Law and Human Behavior.” *Northwestern University Law Review* 97:1115–63.
- Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108(4):814–34.
- Haidt, J., and F. Bjorklund. 2007. Social intuitionists answer six questions about morality. Pp. 181–217 in *Moral Psychology*, edited by Sinnott-Armstrong, W. Cambridge MA: MIT Press.
- Haidt, J., and C. Joseph. 2004. Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus* 133(4):55–66.
- Haisley, E. C., and R. A. Weber. 2010. Self-serving interpretations of ambiguity in other-regarding behavior. *Games and Economic Behavior* 68(2):614–25.

- Halali, E., Y. Feldman, and M. Banaji (in preparation). Enhancing the objectivity of “good” people in subtle conflict of interest situations
- Hollis, J. 2008. *Why Good People Do Bad Things: Understanding our Darker Selves*. New York: Gotham Books.
- Hsee, C. K. 1995. Elastic justification: How tempting but task-irrelevant factors influence decisions. *Organizational Behavior and Human Decision Processes* 62(3):330–37.
- Jolls, C. 2007. Behavioral law and economics. National Bureau of Economic Research Working Paper no.12879. Available at <http://www.nber.org/papers/w12879>.
- Jolls, C., C. R. Sunstein, and R. Thaler. 1998. A behavioral approach to law and economics. *Stanford Law Review* 50:1471–550.
- Jolls, C., and C. R. Sunstein. 2006. Debiasing Through Law. *Journal of Legal Studies* 35:199–241.
- Jones, J. T., B. W. Pelham, M. Carvallo, and M. C. Mirenberg. 2004. How do I love thee? Let me count the Js: Implicit egotism and interpersonal attraction. *Journal of Personality and Social Psychology* 84:5 665–83.
- Jones, J. T., B. W. Pelham, M. C. Mirenberg, and J. J. Hetts. 2002. Name letter preferences are not merely mere exposure: Implicit egotism as self-regulation. *Journal of Experimental Social Psychology* 38:170–77.
- Kahan, D. M. 1997. Ignorance of law is an excuse—but only for the virtuous. *Michigan Law Review* 96:127–54.
- Kahan, D. 2013. Ideology, motivated reasoning, and cognitive reflection: An experimental study. *Judgment and Decision Making* 8: 407–24. Yale Cultural Cognition Lab Working Paper No. 107. Available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2182588](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2182588).
- Kahan, D. M., and E. A. Posner. 1999. Shaming White-Collar Criminals: A Proposal for Reform of the Federal Sentencing Guidelines. *The Journal of Law and Economics* 42(S1):365–92.
- Kahneman, D. 2011. *Thinking, Fast and Slow*. London: Macmillan.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1986. Fairness and the assumptions of economics. *Journal of Business* 59(4):285–300.
- Kawakami, K., J. F. Dovidio, and A. Dijksterhuis. 2003. Effect of social category priming on personal attitudes. *Psychological Science* 14(4):315–19.
- Kay, A. C., S. C. Wheeler, J. A. Bargh, and L. Ross. 2004. Material priming: The influence of mundane physical objects on situational construal and competitive behavioral choice. *Organizational Behavior and Human Decision Processes* 95(1):83–96.
- Kern, M. C., and D. Chugh. 2009. Bounded ethicality: The perils of loss framing. *Psychological Science* 20(3):378–84.
- Kidder, R. M. 2009. *How Good People Make Tough Choices: Resolving the Dilemmas of Ethical Living*. New York: HarperCollins.
- Kish-Gephart, J. J., D. A. Harrison, and L. K. Treviño. 2010. Bad apples, bad cases, and bad evidence about sources of unethical decisions at work. *Journal of Applied Psychology* 95:1–31.
- Korobkin, R., and T. S. Ulen. 2000. Law and behavioral science: Removing the rationality assumption from law and economics. *California Law Review* 88:1051–144.
- Kouchaki, M., F. Gino, and A. Jami. 2014. The burden of guilt: Heavy backpacks, light snacks, and enhanced morality. *Journal of Experimental Psychology: General* 143(1): 414–24.
- Kouchaki, M., K. Smith-Crowe, A. P. Brief, and C. Sousa. 2013. Seeing green: Mere exposure to money triggers a business decision frame and unethical outcomes. *Organizational Behavior and Human Decision Processes* 121(1): 53–61.

- Krieger, L. H. 1995. The content of our categories: A cognitive bias approach to discrimination and equal employment opportunity. *Stanford Law Review* 47(6):1161–248.
- Kunda, Z. 1987. Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology* 53(4):636–47.
- Kunda, Z. 1990. The case for motivated reasoning. *Psychological bulletin* 108(3):480–98.
- Langevoort, D. C. 1998. Behavioral theories of judgment and decision making in legal scholarship: A literature review. *Vanderbilt Law Review* 51:241–96.
- Logue, K. D. 2006. Optimal tax compliance and penalties when the law is uncertain. Michigan Law & Economics Working Paper no. 06-009.
- Lobel, O., and O. Amir. 2012. Risk Management for the Future: Age, Risk, and Choice Architecture. *7th Annual Conference on Empirical Legal Studies Paper*.
- Marquardt, N., and R. Hoeger. 2009. The effect of implicit moral attitudes on managerial decision-making: An implicit social cognition approach. *Journal of Business Ethics* 85(2):157–71.
- Mazar, N., O. Amir, and D. Ariely. 2008. The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research* 45(6):633–44.
- Merritt, A. C., D. A. Effron, and B. Monin. 2010. Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass* 4(5):344–57.
- Mikulincer M., and P. R. Shaver, eds. *The Social Psychology of Morality: Exploring the Causes of Good and Evil*. Washington, DC: American Psychological Association.
- Monin, B., and D. T. Miller. 2001. Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology* 81(1):33–43.
- Moore, C., J. R. Detert, L. K. Treviño, V. L. Baker, and D. M. Mayer. 2012. Why employees do bad things: Moral disengagement and unethical organizational behavior. *Personnel Psychology* 65(1):1–48.
- Moore, D. A., and G. Loewenstein. 2004. Self-interest, automaticity, and the psychology of conflict of interest. *Social Justice Research* 17(2):189–202.
- Moore, D. A., L. Tanlu, and M. H. Bazerman. 2010. Conflict of interest and the intrusion of bias. *Judgment and Decision Making* 5(1):37–53.
- Moore, D. A., P. E. Tetlock, L. Tanlu, and M. H. Bazerman. 2006. Conflicts of interest and the case of auditor independence: Moral seduction and strategic issue cycling. *Academy of Management Review* 31(1):10–29.
- Niedenthal, P. M., L. W. Barsalou, P. Winkielman, S. Krauth-Gruber, and F. Ric. 2005. Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review* 9(3):184–211.
- Nuttin, J. 1987. Affective consequences of mere ownership: The name letter effect in twelve European languages. *European Journal of Social Psychology* 17(4):381–402.
- Pillutla, M. M. 2011. When Good People Do Wrong: Morality, Social Identity, and Ethical Behavior Pp. 353–370 in *Social Psychology and Organizations*, edited by D. De Cremer, R. van Dijk, and J. K. Murnighan. Routledge, New York.
- Pizarro, D. A., and P. Bloom. 2003. The intelligence of the moral intuitions: A comment on Haidt (2001). *Psychological Review* 110:197–98.
- Rand, D. G., J. D. Greene, and M. A. Nowak. 2012. Spontaneous giving and calculated greed. *Nature* 489(7416):427–30.
- Sachdeva, S., R. Iliev, and D. L. Medin. 2009. Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science* 20(4):523–28.



- Schnall, please see p. 5 S., J. Benton, and S. Harvey. 2008. With a clean conscience cleanliness reduces the severity of moral judgments. *Psychological Science* 19(12):1219–22.
- Schurr, A., I. Ritov, Y. Kareev, and J. Avrahami. 2012. Is that the answer you had in mind? The effect of perspective on unethical behavior. *Judgment and Decision Making* 7: 679–88.
- Shalvi, S., J. Dana, M. J. J. Handgraaf, and C. K. W. De Dreu. 2011. Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes* 115(2):181–90.
- Shalvi, S., O. Eldar, and Y. Bereby-Meyer. 2012. Honesty requires time (and lack of justifications). *Psychological Science* 23:1264–70.
- Shoben, E. W. 1983. Use of statistics to prove intentional employment discrimination. *Law and Contemporary Problems* 46:221–45.
- Shu, L. L., F. Gino, and M. H. Bazerman. 2011. Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin* 37(3):330–49.
- Stanovich, K. E., and R. F. West. 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23(5):645–65.
- Stout, L. 2010. *Cultivating conscience: How Good Laws Make Good People*. Princeton, NJ: Princeton University Press.
- Tenbrunsel, A. E., K. Diekmann, K. Wade-Benzoni, and M. Bazerman. 2007. The ethical mirage: A temporal explanation as to why we aren't as ethical as we think we are. Harvard Business School Working Paper no. 08-012. Available at <http://www.hbs.edu/faculty/Publication%20Files/08-012.pdf>.
- Tenbrunsel, A. E., and D. M. Messick. 2004. Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research* 17(2):223–36.
- Tetlock, P. E., and J. S. Lerner. 1999. The social contingency model: Identifying empirical and normative boundary conditions for the error-and-bias portrait of human nature. Pp 571–85 in *Dual-process Theories in Social Psychology*, edited by S. Chaiken and Y. Trope. New York: Guilford Press.
- Thompson, L., and G. Loewenstein. 1992. Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes* 51(2):176–97.
- Trevino, L. K. 1986. Ethical decision making in organizations: A person-situation interactionist model. *Academy of management Review* 3:601–17.
- Uhlmann, E. L., D. A. Pizarro, D. Tannenbaum, and P. H. Ditto. 2009. The motivated use of moral principles. *Judgment and Decision Making* 4(6):476–91.
- Van den Bos, K., S. L. Peters, D. R. Bobocel, and J. F. Ybema. 2006. On preferences and doing the right thing: Satisfaction with advantageous inequity when cognitive processing is limited. *Journal of Experimental Social Psychology* 42(3):273–89.
- Vohs, K. D., N. L. Mead, and M. R. Goode. 2006. The psychological consequences of money. *Science* 314(5802):1154–56.
- Vohs, K. D., N. L. Mead, and M. R. Goode. 2008. Merely activating the concept of money changes personal and interpersonal behavior. *Current Directions in Psychological Science* 17(3):208–12.
- Von Hippel, W., and R. Trivers. 2011. The evolution and psychology of self-deception [target article and commentary]. *Behavioral & Brain Sciences* 34:1–56.



- Yang, Q., X. Wu, X. Zhou, N. L. Mead, K. D. Vohs, and R. F. Baumeister. 2013. Diverging effects of clean versus dirty money on attitudes, values, and interpersonal behavior. *Journal of Personality and Social Psychology* 104(3):473–89.
- Zhong, C. B., and K. Liljenquist. 2006. Washing away your sins: Threatened morality and physical cleansing. *Science* 313(5792):1451–52.
- Zimbardo, Philip. 2007. *The Lucifer Effect: Understanding How Good People Turn Evil*. New York.

## CHAPTER 10

---

# LAW, MORAL ATTITUDES, AND BEHAVIORAL CHANGE

---

KENWORTHEY BILZ AND JANICE NADLER

### 1 INTRODUCTION

---

CLASSICALLY, the ambition of legal regulation is to change behaviors. Laws might aim to increase or decrease various activities, such as owning a gun, or taking a work leave to care for a sick family member, or polluting, or hiring a minority job candidate. They might aim to get people or institutions to substitute one activity for another, such as buying diet soda instead of sugared, or using chewing tobacco instead of smoking, or using solar energy instead of conventional sources. Legal regulation can accomplish its goals directly, through fear of sanctions or desire for rewards. But it can also do so indirectly, by changing *attitudes* about the regulated behaviors. Ironically, this indirect path can be the most efficient one, particularly if the regulation changes attitudes about the underlying *morality* of the behaviors. This is because if laws change moral attitudes, we reduce—maybe drastically—the need for the state to act on or even monitor regulated players.

What does regulation designed to affect moral attitudes look like? It can be obvious, such as an information campaign. But less obvious (and even unintentional) approaches are more common, and probably more effective. Legal regulation might seek to link a behavior that the public already thinks is bad or objectionable to a behavior it currently finds inoffensive. For instance, regulators might wish to reduce abortions generally by stigmatizing a rarely performed, but particularly distasteful, version of it (“partial birth” abortions), with the idea that highlighting—and outlawing—a repugnant version of the procedure will ineluctably associate it with the more common, and more innocuous one. Or legal regulation might recharacterize behavior generally thought of as harmless, or as harmful only to the one who engages in it, into behavior with costs. The regulation of cigarettes is a good example of this phenomenon (Lessig 1995). Or, regulation might tax or subsidize with the hope not only of immediately changing the frequency of behavior, but also

either normalizing or demonizing it. Cigarettes (and “sin taxes” generally) are examples of law imposing a tax on behavior to both discourage and discredit it. On the other hand, law can both encourage and signal approval. For example, to encourage breastfeeding, the law could mandate accommodations that normalize the behavior as natural, and not a shameful activity to be hidden away. In line with this, in the United States, Section 207 of the Patient Protection and Affordable Care Act (29 U.S.C.A. § 207 (West 2011)) requires employers to permit employees who are nursing mothers to take breaks to express breast milk, in a place other than a bathroom.

Any one of these methods could either succeed or fail. They can also have unintended and even perverse effects, including backlashes. The aim of this chapter is to sketch out a framework for understanding the conditions under which regulation succeeds or fails to change underlying attitudes. To begin this chapter, consider two brief case studies in the United States, each well funded and ambitious examples of attempts to change moral attitudes. One was successful (sexual harassment) and another was not (gun control).

## 1.1 Sexual Harassment

In 2012, a unanimous Iowa Supreme Court affirmed the dismissal of a gender discrimination case (*Nelson v. James H. Knight DDS, P.C.* 2012). The complaint was brought by a dental hygienist, who argued that she was fired because her boss found her attractive, and his wife was jealous. Here, the employee was not claiming that she was being coerced into a sexual relationship she did not want, nor that she was being offered a quid pro quo for providing her boss sexual favors, nor that the workplace or her treatment there constituted an environment hostile to women generally. Any one of these assertions would have constituted a face-credible claim for sexual harassment,<sup>1</sup> and she would not likely have lost on summary judgment.

Her claim was different: she argued that *but for* her being female, she would not have been fired—because if she were not female, her boss would not have found her sexually desirable and his wife would not have been jealous of her. Unfortunately for the plaintiff, the law of sexual harassment is quite settled here—it is not enough for a plaintiff to prove that but for her gender she would not have been fired. She must prove that her gender itself *motivated* the firing—and sexual jealousy by the spouse of the employer is a legally permissible motive distinct from the employee’s gender.<sup>2</sup>

The legal resolution of this case was quite predictable. What was *not* predictable was the public reaction. The decision attracted national attention, and almost

<sup>1</sup> The federal law governing discrimination in the workplace is Title VII of the Civil Rights Act of 1964. The Iowa equivalent, under which the dental hygienist filed her own case, is the Iowa Civil Rights Act, Iowa Code 216.6(1)(a). Iowa courts turn to federal law when interpreting their own discrimination cases, *Deboom v. Raining Rose, Inc.*, 772 N.W.2d 1, 7 (Iowa 2009).

<sup>2</sup> See, for example, *Tenge v. Phillips Modern Ag Co.*, 446 F.3d 903 (8th Cir. 2006); *Platner v. Cash & Thomas Contractors, Inc.*, 908 F.2d 902 (11th Cir. 1990).

uniformly, the decision was lambasted and ridiculed (Newcomb 2012; Politi 2012). Many news outlets noted within the first sentence or two that the Iowa court was “all male,” and that the court had used the term “irresistible” to describe the plaintiff (Foley 2012; Associated Press 2012). Countless bloggers and commentators derided the decision, in often colorful terms; editorialists have even called for reform to close this “loophole” in the law of employment discrimination (Cassens Weiss 2013; Fiedler 2013).

This overwhelming reaction is a strong sign of the success of antidiscrimination law in shaping the moral attitudes of the general public about sexual harassment. Though “Title VII and the Iowa Civil Rights Act are not general fairness laws, and an employer does not violate them by treating an employee unfairly so long as the employer does not engage in discrimination based upon the employee’s protected status” (*Nelson* at p. 9), citizens’ expectations about how they are entitled to be treated at work have clearly changed over the last three decades, ever since courts began recognizing sexual harassment as discrimination. Arguably, Title VII played a *causal* role in this process (Sunstein 1996). It is hard to imagine op-eds like the following being published by the right-leaning *Chicago Tribune* in the seventies, the eighties, or possibly even the nineties:

If Knight wasn’t guilty of illegal discrimination, he was certainly guilty of being a creep and a boor—not to mention a tightwad, having rewarded Nelson’s years of stellar work with a single month’s severance pay. About the best that can be hoped is that the publicity surrounding the poor treatment Nelson received will attract the interest of dentists at another practice who would place a high value on her excellent skills and performance—and pay no attention to her looks. And Knight? Maybe he needs to find a new place to apply his Novocain. (“You’re Irresistible. You’re Fired.” 2012)

A woman working outside the home in the sixties would have had to expect the possibility of “flirtation” (or predation) as a condition of employment, plus strongly gendered expectations about appearance and the propriety of romantic relationships in the office (Farley 1980). But today, the norms of professionalism have moved towards not just equality, but altogether away from such sexualization in the workplace. Simply put, when evaluated in terms of shifting public attitudes, sexual harassment law is an almost unvarnished success. This does not mean sexual harassment has been eliminated, of course—but it has certainly been marginalized. Behavior and attitudes that would have been not only acceptable, but normal, just a couple of decades ago, are now anathema.

## 1.2 Gun Control

Not so, attitudes about guns. The history of gun control in the United States is long and fascinating, if frequently depressing. From the Civil War until the 1970s, gun control activism often was motivated by a desire to limit access to guns by African

Americans (Winkler 2011). Over this period, attitudes about guns have remained remarkably stable. A 1972 empirical review of various public opinions polls from the 1930s until the article's publication revealed that a relatively consistent three-quarters of Americans have supported limited forms of gun control, such as registration (Erskine 1972).<sup>3</sup> More recent reviews have shown the same: around three-quarters support regulations like regulation and background checks (Smith 2007; Stinchcombe 1980). Opposition to making laws covering the sale of firearms more strict has been stable as well (Gallup 2012).

This is striking, given the pattern of gun laws tightening and then loosening over the past century in the United States (Kopel 2012). In other words, activists on both sides of the gun control debate have invested decades of time, and millions of dollars, investing in lobbying efforts to change Americans' relationship to guns. They have something to show for their efforts, in that gun *laws* have changed quite a bit, in both directions, over the last one hundred years. Empirical evidence suggests that lobbying activities do directly influence whether particular gun laws get passed (Langbein and Lotwis 1990; Langbein 1993), but there is little evidence that gun *attitudes* have changed in response to interest groups' efforts. One targeted study examined attitudes about a particular proposed reform in Virginia, and found that public attitudes in the period before the legislative vote were the same as those measured by a different pollster six months after the legislative vote was done (Kauder 1993). Though there are spikes in support for gun control policies in the wake of highly publicized shootings, attitudes quickly revert (Newport 2012). Given the prominence of the gun debate, and the stakes each side asserts, this equilibration is astonishing.

Simply put, we argue in this chapter that laws themselves sometimes seem to affect moral attitudes in the intended ways, and sometimes not. Can we make any generalizations—or better yet, predictions—about when it will do which? We think so. We argue that the success of legal regulation in changing moral attitudes will depend on a number of variables. The ones we focus on in this chapter are (1) whether the regulation aims to change attitudes which are important to individuals' cultural *identities*; (2) whether there is underlying *dissensus* about the behavior or attitude; and (3) whether the law is attempting to change the underlying *meaning* of behaviors rather than trying to change the behaviors itself.

We begin by briefly reviewing the literature on the influence of law and legal institutions as a whole on moral attitudes and behaviors. We examine the idea that the legal system can promote compliance with law and cooperation with legal actors and institutions when it is perceived as legitimate and furthering the aims of justice. We then turn to the ability of specific laws and legal regulations to influence moral attitudes and behaviors. We examine the influence of law through various mechanisms, including physical architecture, social meaning, attitude change, and consensus. Throughout

<sup>3</sup> The percentages fluctuated from a brief high of 84% in one poll in 1969 to an equally brief low of 66% in 1971 (Erskine 1972).

the discussion of these mechanisms, we focus on factors that lead to success, failure, or even perverse effects.

## 2 GENERAL LEGITIMACY AND CREDIBILITY OF THE LEGAL SYSTEM

---

Undoubtedly, laws are sometimes effective because they are backed by the threat of punitive enforcement. This threat prompts individuals to make judgments about risk and reward before deciding whether to engage in a prohibited activity. Accordingly, standard economic analysis relies on the assumption that if the expected cost of behavior—comprised of the severity and probability of punishment—exceeds its expected benefit, then the actor will refrain from that behavior (Becker 1968). Indeed, standard economic analysis assumes that questions about the effect of law on human behavior both begin and end with the assumption that behavior responds to rewards and punishments. At the same time and in parallel, law forthrightly attempts to shape citizens' moral beliefs. When the law forbids murder, this is because murder is evil, and the language of the law sometimes makes explicit the moral implications of the prohibited act. Thus, the criminal law traditionally grades an unintentional killing committed with a “depraved heart” or an “abandoned and malignant heart” as murder, but other unintentional killings as merely manslaughter. Conversely, Good Samaritan laws—although uncommon in the United States—are designed to shape beliefs about the moral duty to rescue. In this section, we examine the claim that law can succeed in changing moral attitudes through its general reputation for doing justice. We very briefly explore two literatures that consider ways in which law's general reputation works to guide moral attitudes, and ultimately, behaviors governed by those attitudes. Specifically, we first examine the argument that people defer to law because its content is deemed generally consonant with popular morality. We then examine the argument that legal agents promote cooperation by enacting principles of procedural justice.

### 2.1 Legal Content as a Reliable Source of Morality

To the extent that the legal system is perceived as promoting justice, people will be more likely to comply with law in an overarching sense (Robinson and Darley 1996; Robinson and Darley 2007; Nadler 2005). Compliance with law can mean doing what is required by law such as paying taxes or picking up dog waste, or it can mean refraining from prohibited acts like discriminatory employment decisions or vigilante violence against wrongdoers. In general, when the law imposes obligations and punishment in concordance with general intuitions about justice, then people are more likely to view the legal system as a legitimate and reliable source of morality. Individual cases decided

consistently with lay intuitions of justice reinforce the notion that the law is a source of moral guidance. Discrete laws that reflect the community's sense of justice will enhance the ability of the law to gain compliance in people's everyday lives, in areas unrelated to the law in question (Mullen and Nadler 2008).

The perceived legitimacy of the law can be undercut, and compliance undermined, when the law fails to comport with citizens' intuitions of justice. The mass incarceration of young African American males within communities, for example, is perceived by many as evidence of the law imposing punishment in an unjust fashion (Alexander 2012). One possible consequence is that community members who perceive the law as unjust are less likely to comply with the law themselves (Mullen and Nadler 2008; Nadler 2005). Disaffected citizens might also resort to vigilantism, thereby opting out of a legal system they perceive as deviating from the consensus of the community. When there is community consensus about what justice requires of law, then it is fairly straightforward, at least in theory, to implement this consensus and enact justice. When the legal system comports with justice, it gains legitimacy and compliance because citizens look to law as a source of moral guidance.

A more difficult situation arises when there is no clear consensus about what justice requires, or worse, when there is clear dissensus. For example, no matter what the position the law takes regarding abortion, a large contingent of citizens will believe the law to be wrong and even immoral. Some individuals have a strong moral investment in either permitting or prohibiting abortion, and for them abortion laws that permit immoral outcomes can prompt strident protests and even vigilante action. For example, in the United States, extremists who believe that the government abortion laws sanction the murder of babies have bombed abortion clinics and murdered physicians (Darley, Tyler, and Bilz 2003). Other examples of dissensus abound. People's intuitions differ, often along predictable cultural lines, about the morality of severe punishment for abused women who kill their sleeping husbands, a man who kills another man because he solicited sex, the possession of recreational drugs, among many other examples (Braman, Kahan, and Hoffman 2010).

## **2.2 Legitimate Legal Actors and Institutions Providing Motivation for Cooperation**

Perceptions of the legal system are shaped by people's everyday experiences with law through interactions with institutions such as the police, traffic court, family court, and government agencies. Even people with no direct personal experience with the legal system nonetheless can readily observe the pervasive operation of legal institutions and authorities. The ways in which people experience the operation of law in their everyday lives gives rise to perceptions about the legitimacy of the law and its actors and institutions. To the extent that people perceive legal authorities to be legitimate, they will view those authorities as properly empowered to make decisions about how the law should be implemented (Tyler and Huo 2002; Tyler 2006a). They will not readily



question when they see a police officer making a traffic stop or when they read about a judge who approves a large employment discrimination award. But to the extent that the opposite is true—that is, when people believe legal authorities lack legitimacy—they will view authorities as exercising power improperly (Tyler 2009).

Empirically, perceptions of legitimacy turn on the quality of treatment people receive from and observe on the part of legal actors and institutions (Tyler and Fagan 2008). Examples of poor quality of treatment by legal actors and institutions include judges who convey a lack of concern toward citizens who appear in their courtroom as litigants or jurors; police officers who convey an attitude of disrespect toward people they stop or question; police department policies and practices that signal surveillance, suspicion, and mistrust, leading people to feel that their communities are being harassed rather than served. These types of perceptions determine the extent to which people see the actions of judges and police as moral and legitimate (Tyler and Wakslak 2004; Tyler 2006b). Legal authorities and institutions that are perceived as legitimate are likely to prompt community members to perceive cooperation and compliance as moral imperatives. Conversely, legal authorities and institutions lacking legitimacy are less likely to prompt people to feel morally obligated to cooperate, and thus less likely to take voluntary actions like calling 911 to report suspicious activity; coming to the police station to identify a criminal; coming to court to testify; accepting the decision of a family court regarding custody; showing up for jury duty and doing a conscientious job when serving; paying fines for traffic tickets; paying income taxes and declaring all income. In sum, a legal system that lacks legitimacy produces community members less likely to feel morally obligated to defer to legal decisions and to cooperate with legal actors and institutions (Tyler 2006b; Tyler and Huo 2002).

### 3 MECHANISMS OF INFLUENCE

---

In this section, we review how specific attempts to regulate can succeed or fail to bring about change in people's moral attitude about the behavior in question. Previous scholarship on expressive law has emphasized that law can express values, which in turn can influence behavior (see, e.g., Cooter 1998; McAdams 2000a; Sunstein 1996). Some of the prior research on expressive law explores various mechanisms by which law influences behavior expressively, such as by leveraging people's motivation to maintain the esteem of others (McAdams 1997), or by providing a focal point in situations where coordination is required (McAdams 2000b). We begin by reviewing how law can use physical architecture to change behavior and possibly moral attitudes, by making the targeted behavior either more convenient (e.g., recycling) or less convenient (e.g., smoking). We then consider how legislation or regulation can change the social meaning of a behavior, rendering what was formerly considered to be moral or amoral to be now morally problematic (e.g., discrimination), or vice versa (homosexuality). Next, we briefly review some classic literature in the social psychology of attitude change to highlight

factors that make it less likely that law will successfully influence moral attitudes. We also consider the effects of perceived democratic consensus of law on moral attitudes.

### 3.1 Architectural Nudges

Bernstein uses the term “melioristic law reform” to refer to citizen-driven efforts to change tort law and criminal law in an effort to improve individual behavior (Bernstein 1996). Individuals sometimes change their behavior readily if gently nudged in a particular direction (Kahan 2000; Sunstein 1996). The most promising domains for changing behavior are those in which people can be subtly prompted to take the path of least resistance. If law makes a desirable behavior—such as recycling—more convenient, people can be persuaded to do more of it. If law makes an undesirable behavior—such as smoking—less convenient, people can be persuaded to do less of it. Initially, the changes are solely behavioral, and up to this point, conventional law and economics can explain changes in people’s willingness to engage in the regulated behavior. But along the way, attitudes about the morality of the behavior can get altered as well. The propensity of behavior to influence attitudes has been long known in social psychology—this is the mechanism underlying classic demonstrations of cognitive dissonance. For example, people asked to choose between two closely ranked products later increase their liking of the chosen item and decrease their liking of the forgone item (Brehm 1956). For attitude change to follow, however, it is important that the individual perceive the behavior as a product of her own choice (Cooper and Fazio 1984). This is another departure from the lessons of conventional law and economics—if the regulatory “nudge” is so forceful that any behavioral changes are seen as obviously caused by legal incentives rather than by free will, attitude change is unlikely to occur (Kuran 1995). Similarly, although in some circumstances attitude change can be assisted by public information campaigns, they are an unlikely source of moral attitude change, as we discuss later.

As an example of the use of architectural means to prompt behavioral and attitudinal changes, consider the case of smoking in the United States. In 1964 the US surgeon general’s report established the scientific basis supporting the claim that smoking carries grave health risks for smokers. The following year, Congress mandated health warnings on cigarette packages and in 1969 imposed limits on tobacco advertising (Federal Cigarette Labeling and Advertising Act, 15 U.S.C. §§ 1331–40; Federal Public Health Cigarette Smoking Act of 1969, 15 U.S.C. §§ 1331–40). It is unclear whether this type of government regulation of information—health warnings and advertising limits—caused any discernible change in smoking behavior. But a turning point came when the prospect emerged that nonsmokers were being harmed by secondhand smoke. At that point, smoking became more than just a question of an individual smoker’s liberty interest in engaging in an activity that causes harm to the self—now harm to others was an issue (Jacobson, Wasserman, and Anderson 1997). Because nonsmokers were being subjected to harms to their health that they did not always consent to, the issue took on

a more salient moral dimension, and legal restrictions on smoking behavior became a real possibility. By the early 1970s, laws imposing bans on smoking in certain public places began to slowly proliferate. In 1986, the US surgeon general released a report that found that secondhand smoke causes cancer, and later reports found that secondhand smoke causes lung and heart disease (“The Health Consequences of Involuntary Smoking” 1986). In children, secondhand smoke is associated with asthma, ear infections, and respiratory infections (“The Respiratory Health Effects of Passive Smoking” 1992). Conventional economics would predict that the largest effects on smoking behavior would stem from effects on the smoker herself. Instead, as elaborated below, the largest effects emerged from realizations about its effects on *others*.

Following the release of these reports, local smoking bans proliferated in some parts of the United States followed by statewide bans in some states (Jacobson, Wasserman, and Anderson 1997). California enacted strong statewide smoking regulations in 1994 that banned smoking in offices, and later restaurants (Cal. Lab. Code § 6404.5 (Act of July 21, 1994, § 1, 1994 Cal. Stat. 310)). In New York, local bans were followed by a strict statewide regulation in 2003 that prohibited smoking in all restaurants (N.Y. Pub. Health Law 1399-n to -x). In many localities, smoking is now prohibited even in prototypical places for smoking like bars and casinos (see e.g., Colo. Rev. Stat. § 25-14-204 (casinos); Smoke Free Illinois Act, 410 ILCS 82 (bars, restaurants, and casinos); N.Y. Pub. Health Law 1399-o (bars and restaurants); Cal. Lab. Code § 6404.5(f)(1)(a) (bars and restaurants)). Arguably, the political will that made these state and local regulations possible was a result of several things; here, we focus on two. The first is the notion that smoking harms others besides the smoker, so that the smoker is imposing a real cost on others against their will, and therefore others have the right to object. The second is the increase in state and local taxes on cigarette purchases, which helped boost the fortitude of those who wanted to quit smoking. As to the first factor, by prohibiting smoking from most public places, legal regulation shapes behavior directly, by leveraging inertia. That is, when smoking is made difficult, it forces the smoker to wait to smoke, to get up and move to a different location, possibly into bad weather; as a result, smokers are likely to smoke less often. Further, those smokers already considering quitting—perhaps because of fear of dreaded disease—are further encouraged to do so. As to the second factor, increased cigarette prices obviously increase the will of smokers to quit. But interestingly, the justification for the taxes also played on the third-party effects of smoking—they were promoted as at least partially offsetting the public costs of chronic illnesses caused by smoking (Santora 2008).

In addition to making use of architecture to discourage harmful behavior, law can also encourage prosocial behavior. Consider recycling. As late as the early 1990s, many communities in the United States, Canada, and Europe did not have mandatory recycling and curbside pickup programs. Municipal regulation that mandated curbside recycling pickup increased dramatically in this period, and the availability in the United States of curbside pickup of recycled materials was partly responsible for an enormous increase in the amount of solid waste that was recycled (Kinnaman 2006). Local regulation making recycling easier and more convenient (by, for instance,

picking it up at curbside along with the regular trash) encourages the target behavior, even on a conventional law and economics analysis. But the removal of small barriers can have disproportionately large effects. For example, separating items probably does not take more than a total of a minute or two of time per day. But eliminating this barrier increases recycling dramatically (Carlson 2001). It is interesting to note that unlike in the example of smoking regulations, the regulations that provide the infrastructure for recycling do not involve prohibiting or requiring anything on the part of citizens. Yet even in the absence of mandatory recycling, making it easier makes it more prevalent. Once recycling becomes a behavioral regularity, positive attitudes toward recycling follow, especially if the individual perceives her own commitment to recycling as voluntary. When an individual perceives herself as the kind of person who recycles, that self-concept leads to valuing the idea of recycling (Werner et al. 1995; Cialdini 2009).

Law reduced smoking by erecting barriers and increased recycling by removing barriers. Again, one way law encourages prosocial behavior is through positive incentives. For example, in Italy, national law provides that any employee who donates blood is permitted to take one entire day off work with pay, per year (D.L. 21 Ottobre 2005, n. 219, in G.U. October 27, 2005, n. 251(It)). This law induces employees to make an average of one extra blood donation per year—a seemingly straightforward response to incentives, and an example of how law can accomplish its goals directly through fear of sanctions or desire for rewards. But possibly, providing a day off of work for blood donation also changes moral attitudes about blood donation. Specifically, when workers begin donating blood in exchange for the day off, the behavior becomes regularized, and people continue to donate at the same frequency even when they stop being employed. And indeed, the facts bear this out: unemployed donors who later become employees make about one extra donation per year, while there is almost no reduction in donation frequency by donors who cease to be employees—and who thereby cease to receive the day-off benefit. Thus, people generally do not stop providing their extra donation when they no longer receive the benefit the law provides. The donation behavior incentivized by law appears to be transformed into a habit, and possibly even internalized as morally good (Lacetera and Macis 2012; Werner et al. 1995).

### 3.2 Social Meaning Changes

The social meaning of any given behavior depends on the existing social norms governing that behavior (Sunstein 1996). Architectural changes that make a behavior less prevalent can also contribute to changes in the social meaning of that behavior. Consider again the example of smoking. Aside from physical and temporal inconvenience, legal regulation limiting smoking also imposes a social cost on smokers. Being cast out of offices, restaurants, and even bars sends a message of banishment and ostracism to the smoker, which for some might be internalized as shame. In order to engage in the behavior that the law regulates, smokers are increasingly confined to their homes

or designated smoking zones, and smoking is increasingly perceived as an act of social deviance, at least within some social circles or in some times and places. When the issue was one of the individual smoker's decision to harm herself, legal regulation was politically difficult because of the strong libertarian undercurrent that often characterizes policy debates in the United States about hazardous activities. On this view, smoking is a decision about personal preference, like eating chocolate or wearing shorts. Rozin and Singh (1999) argue that over the course of the latter half of the twentieth century, smoking evolved from being viewed as solely a matter of personal preference without implications for moral worth, to being an object of disgust, which empowered nonsmokers to condemn the activity. In societies like the United States which are marked by cultural individualism, an activity like smoking is not psychologically endowed with moral status unless it can be demonstrated that it causes harm to others (Schweder et al. 1997). In this way, the findings that secondhand smoke harms others, especially children, set the stage for bringing the issue of smoking into the domain of the moral, making it politically easier to regulate through law (Rozin 1999). As Dan Kahan has argued, information about the hazards of smoking was accepted and believed only after a cultural shift in the social meaning of smoking as morally noxious because of the dangers to nonsmokers (Kahan 2007). Tellingly, now even other nicotine delivery systems that do not impose harms on others are being regulated because of their association with the "tainted" activity of smoking. That is, legal regulation of one activity can not only change the meaning of that activity, but can spread to others, too. For instance, California has proposed a ban on using increasingly popular "e-cigarettes" in public places, which produce only water vapor (and possibly trace amounts of nicotine) instead of smoke (Cal. Sen. Bill 648, Feb. 22, 2013). At least one municipality has forbidden its employees from using smokeless tobacco while at work. The town manager told the city council, "The concern is a number of employees are chewing tobacco in the presence of a vehicle and others, and they find it very disturbing and they would like to ban the use of chewing tobacco. People are just offended by tobacco products. They think it's a filthy, vile habit and they resent other employees doing it" (Graff 2013).

Bringing an activity out of the domain of the nonmoral and into the domain of the moral can increase prosocial behavior, just as it can decrease the incidence of harmful behavior. Consider again the example of recycling. If people perceive that their neighbors are recycling, it is more likely that they will recycle too. Just as smokers might be motivated to quit if they feel banished and shunned by regulations on smoking, householders might be motivated to recycle if doing so makes them feel included and virtuous within their neighborhood. In an early study of residents of a city that had just begun a curbside recycling program, 95% of those who recycled reported that their friends and neighbors did also; but only 67% of nonrecyclers reported that their friends and neighbors recycled (Oskamp et al. 1991). This finding is consistent with results of public goods games in which a norm of cooperation is triggered by cooperation by others (Carlson 2001). This suggests that physical and cognitive inertia is not the only barrier to increasing the rate of prosocial behaviors like recycling. Making it easier to do so is important, to be sure. But in addition to convenience, another important

determinant of prosocial behavior is the subjective meaning of that behavior. If an individual perceives that others are recycling, then they might perceive recycling as an act of cooperation that is expected by other members of the community. Conversely, if an individual perceives that others are not recycling, then he might perceive recycling as an act of substantial sacrifice (Davidai, Gilovich, and Ross 2012; Cialdini 2009; Kahan 2003).

Using less energy is another prosocial behavior that can be encouraged by appeals to social comparison. Home energy use can be reduced by sending customers “home energy reports” comparing their energy use to that of their neighbors (Ayres, Raseman, and Shih 2012). It is not yet clear exactly how receiving peer comparisons serves to reduce energy consumption. One possibility is that people whose consumption exceeds those of their neighbors feel guilt about contributing to the problem of climate change. If this is true, then it is possible that information campaigns such as the energy reports used in this study can work to change attitudes about the morality of energy conservation. Legal regulation might play a role in this type of moral attitude change by mandating information campaigns of this sort.

Finally, legal regulation that alters the social meaning of a behavior can result in a change in the way the people engaging in the behavior are morally evaluated. In the workplace, women with children are, on average, offered lower starting salaries and are less likely to be hired and promoted because they are perceived as less competent and less committed to their jobs compared to childless women. In 1993, the US Congress enacted the Family and Medical Leave Act (FMLA) (29 U.S.C. §§2601–2654), which provides unpaid job-protected time off for family caretaking as a gender-neutral entitlement. Experimental evidence suggests that making the FMLA salient influences normative judgments of women who take caretaking leave, by changing the meaning of taking time away from work to fulfill caretaking responsibilities. It is possible that law is perceived as representing a broad social consensus about the meaning of caretaking leave, and in turn reduces stereotyped inferences people make when they observe a worker who takes family leave (Albiston and Correll 2011).

Legal regulation can therefore transform the social meaning of behavior, changing people’s perceptions regarding the moral acceptability or desirability of the behavior. But sometimes, social meaning change is difficult to manage through regulation when regulation is perceived as attacking the fundamental identity and status of a discrete cultural group. Consider the example of gun regulation in the United States. Like the regulation of smoking, the regulation of guns is designed to reduce the risk of harm. In the case of smoking, harm reduction strategies were mildly controversial to the extent that they were considered inconvenient and expensive by smokers. Contrast this with attempts to regulate gun ownership. As cultural symbols, guns differ from cigarettes because guns signify membership in an important affinity group. According to the cultural cognition thesis, people who are relatively hierarchical and individualistic in cultural orientation tend to oppose gun regulations because they believe law-abiding citizens would be vulnerable to predation as a result of regulation (Kahan et al. 2007). Guns are associated with hierarchical social roles like father, hunter, and protector.



Guns are also associated with individualistic cultural symbols, such as self-reliance, courage, honor, and prowess (Kahan et al. 2007). As a result, hierarchical individualists tend to be skeptical of risk perceptions that would justify such regulations. People who identify with guns include hunters and other outdoor sports enthusiasts who view gun education as an integral part of rural living (Barringer 2013). This helps to explain why the identity of members of this group is undermined and their status is threatened by gun regulation. Gun regulation is perceived by many as a threat to their way of life and an implicit statement that they do not matter. At the same time, there is no discrete cultural group whose identity hinges on members defining themselves by the absence of guns. Thus, unlike the regulation of smoking, the regulation of guns is unlikely to have straightforward influences in shaping behavior and moral attitudes.

### 3.3 Using Law to Directly Influence Moral Attitudes

The literature in social psychology on attitude change generally is far too immense to summarize here (Bohner and Dickel 2011). Instead we focus on those ways that *legal regulation* in particular may change attitudes. That is, we are holding constant the general source of possible attitude change as being “some legal entity”—be it legislation, an agency action, a court decision, or a plebiscite. These different sources, of course, will themselves have different success rates in different settings—but all, at least, could be categorized in some sense as an “official” legal regulator. So, why might official attempts at attitude change either succeed or fail?

#### 3.3.1 *Effects of the Source*

Although all legal regulators might be seen in some sense as representatives of “the state,” these representatives are not monolithic. Each has a domain in which action will be seen as appropriate, by subject matter, by hierarchy (local, regional, national, international), or by mode of action. Briefly, the same attempt at attitude change will be perceived as more or less legitimate, depending on which legal entity issues the regulation. An entity that attempts to act outside its sphere of legitimate influence will be perceived as a bully, or, just as devastatingly, laughable. Consider the example of numerous municipalities declaring themselves as “Nuclear Free Zones” in the 1980’s (Maclay 1988). These laws were mostly symbolic. In the cities where it might have had real bite because of the presence of institutions doing nuclear research, like Cambridge, Massachusetts, and Palo Alto, California, the regulations failed to pass, and those in which it did pass routinely ignored the ordinances or granted waivers so that the local economies could continue their work uninterrupted (Emmanuel 2012). By now, the number of US. municipalities with such bans exceeds one hundred. We know of no study that has attempted to assess the effect of such declarations on public attitudes towards nuclear proliferation, but we are confident that any such study would find nothing. Editorials and commentary on the ordinances are just as likely to reveal ridicule or embarrassment as support, with the following sentiment being representative:



“[I]f anybody—and I mean ANYBODY—tries to bring a nuclear warhead into [Iowa City’s] limits, they will be facing (I am not making this up) a \$500 fine! With a deterrent like that, you understand how Iowa City has distinguished itself from other small towns in Iowa and managed to stand firm against the onslaught of nuclear weapons” (Brawner 2010).

A municipality might try to change attitudes toward local issues such as, say, recycling or cleaning up after pets or shoveling snow from sidewalks, but if it tries to affect attitudes toward international issues such as nuclear disarmament, its attempts are likely to fail. When an institution steps outside of its expected domain to regulate, the public is likely to regard its actions with skepticism. And attempts to regulate that are seen as institutionally misplaced can backfire. Consider again the 2011 Patient Protection and Affordable Care Act, which the US Justice Department (successfully) attempted to make more constitutionally acceptable by characterizing as a “tax” rather than as a mandate for purchase of health insurance. While recasting the regulation as a tax allowed it to win its battle in the Supreme Court (*Nat’l Fed’n of Indep. Bus. v. Sebelius* 2013), its strategy might have diminished the federal government’s ability to win the larger cultural war over US citizens’ attitudes about the moral obligation to acquire health insurance.

Even when acting within the scope and boundaries of its appropriate domain, regulation may fail if the legal entity is perceived as not credible (Sternthal, Phillips, and Dholakia 1978). One way to lack credibility is to be untrustworthy, and classic experimental work in social psychology has shown that untrustworthy sources have little ability to persuade about factual issues and predictions (Hovland and Weiss 1951). Similarly, a source that is perceived as lacking in knowledge will not be persuasive on factual questions (Kuklinski and Hurley 1994). Still another form of noncredibility has to do with illegitimacy—and this is particularly relevant for failures to persuade on *moral* as opposed to factual issues. Edwin Hollander (1958) argued that leaders build up “idiosyncrasy credits” by conforming their decisions to the expectations of their followers, which they can then “spend” to behave in ways that run counter to expectations. This insight is easily transferred to legal entities rather than to leaders—a legal entity that has built up goodwill may be able to change public attitudes, but one that has been bucking public opinion for years will be relatively helpless to change public attitudes with yet another maverick legal act (Gibson, Caldeira, and Spence 2003). In the wake of more than a decade of decision after unpopular decision establishing rights for criminal defendants in the 1960s, for instance, the Supreme Court arguably simply had no more reputational capital left to spend in changing public attitudes toward school busing (*Swann v. Charlotte-Mecklenburg Board of Education* (1971)). The Court’s popularly excoriated decision led to massive resistance by whites, and even to violence in Boston (Tager 2001).

One researcher found that while attitudes toward busing did change in Charlotte, they did not do so as a direct result of the persuasive effects of the Court’s decision. Instead, attitude change came from their compliance with the Court’s order. That is, after sending their children to integrated schools and finding that the education system

did not collapse as they had expected, cognitive dissonance forces caused them to soften their attitudes toward busing (Maniloff 1979). Timur Kuran (1995) argued that, if the goal is attitude change, the *reason* citizens comply with a legal mandate is less important than *that* they comply and that citizens see others complying. Eventually, citizens assume others are complying not because they must, but because they want to—and this perception contributes to a shift in social norms.

This is also consistent with more modern findings that sources that espouse views that are contrary to the source's usual positions, or which are counter to the source's own or their group's interest, are more influential than those who act in line with expectations. Consider, for example, Republican governor George Ryan's decision to place a moratorium on capital punishment in Illinois. The fact that a Republican governor was an unlikely source of deep skepticism about the death penalty undoubtedly moved the death penalty abolitionist's position forward more than would the same move would have if made by a Democratic governor (Gross and Ellsworth 2001; Visser and Mirabile 2004). Failures caused by source characteristics are more likely to result in simple failures than in backlashes or perverse effects. They will simply be seen as hamhanded, self-interested, incompetent, or silly—and rejected outright by their intended audience.

### 3.3.2 *Effects of the Attitude Object*

If characteristics of the *source* of the regulation can cause it to fail, then so can characteristics of the targeted attitude itself. The most important of these, at least in the United States, is the ease with which we can describe the targeted object of regulation as causing *harm* to third parties. For instance, if we wish to change attitudes about a particular behavior, it helps to characterize the behavior as causing damage not just to those (informed, consenting adults) who choose to engage in it, but to innocent bystanders as well. As discussed earlier, the legal movement against smoking gained traction once the medical risks of "secondhand smoke" were introduced as part of the policy debate. Of course, it is almost trivially simple to describe *any* behavior as having at least some third-party harms (Bilz and Darley 2004). As a natural corollary, then, to the extent opponents can describe those third-party harms as a front, and persuasively argue that the target of attitude change causes mere moral harm (or, perhaps, mere psychic harm), the attempt at legal regulation is likely to fail. In a political society informed by Enlightenment liberalism's emphasis on the Harm Principle, as it is in the United States, citizens often resent attempts by government to impose any kind of official orthodoxy—even when, interestingly, they individually endorse the belief.<sup>4</sup> The strands of individualism in the United States run deep, and an attempt to change moral attitudes *as moral attitudes* is likely to fail (Katyal 2006).

<sup>4</sup> For instance, when asked the following poll question, "Some people think the government should promote traditional values in our society. Others think the government should not favor any particular view. Which comes closer to your own view?" 50% of Americans agreed with the latter statement, and fewer than half with the former (Opinion Research Corporation 2011).

Another element of the attitude object that can undermine successful attempts to change it is how discrepant the legal regulation is from what citizens currently think it should be. The further apart they are, the more resistant to change people will be (Hovland, Harvey, and Sherif 1957). Although in some sense this is obvious, more than one attempted legal regulation has foundered on the shoals of being “too extreme” for the public to accept. Dan Kahan has written about the difference between what he calls “gentle nudges” versus “hard shoves” in legal regulation, arguing that the latter not only tend to fail, but can actually cause backlashes (Kahan 2000) (or, in the terms of early attitude research in social psychology, “boomerang effects” [Sherif, Sherif, and Nebergall 1982]).

Kahan uses the example of date rape. Date rape became a cause célèbre of feminists in the 1980s, especially on college campuses. Activists wished to change the notion that rapes that occurred between acquaintances, where perhaps the issue was confusion about consent (does “no” really always mean “no”?), were somehow less bad than stranger rapes (Viki, Abrams, and Masser 2004). Their solution was to urge that date rape be treated by the legal system and all its actors (police, prosecutors, judges, and juries as fact-finders) as equally bad, by being punished equally (Clemens 1982; Taslitz 2005). Kahan argued that in so doing, activists were actually doing more harm than good. Because most of the public thought date rape was bad, but not *as bad* as stranger rape, insisting that they be treated identically would result in police who were reluctant to arrest, prosecutors who were reluctant to charge, and judges and juries who were reluctant to convict *at all*. This very reluctance would serve as social evidence that, in fact, date rape was not only not as bad as stranger rape, but perhaps was not even really a crime (Kahan 2000). Though research showing the law (or at least, the *attempts* to change the law) directly caused a backlash is lacking here, plenty have argued that the law is impotent to impose a change in attitudes about date rape through dramatic redefinition of the crime (Schulhofer 2000; Pierce Wells 1996), and the phenomenon of “date rape backlash” itself is accepted enough to have spurred considerable commentary and research (Woods and Bower 2001; Clarke, Moran-Ellis, and Sleney 2002) and even a documentary film (Media Education Foundation and FAIR (Firm) 1994).

A final feature of the attitude object that will affect the ability of a law to change attitudes is the level and quality of citizens’ involvement with the issue. The greater the involvement, the more resistant to change people will be (Sherif and Hovland 1961). Social psychological researchers have focused on a number of different ways one can be “too involved” to allow a legal decision-maker to change your mind about something. The issue being discussed could be particularly vivid too frightening to contemplate a change in attitude about it (Leventhal, Meyer, and Nerenz 1980). Or, the issue could have high levels of importance to citizens. Early on, Hovland and Sherif argued that those with high “ego involvement” in an issue would be resistant to persuasion (Sherif and Hovland 1961; Sherif, Sherif, and Nebergall 1982). Much more recently, Linda Skitka has focused on “moral mandates.” Where people’s attitudes are strong, wrapped up in their personal identity, and related to core moral values, they will be particularly resistant to changing them (Skitka and Houston 2001; Mullen and Skitka 2006). Since

this chapter is particularly concerned with changing *moral* attitudes, Skitka's lessons are particularly important. Interestingly, high levels of self-interest play less of a role in determining people's attitudes than standard economic analysis would predict, and so presumably whether or not the proposed legal regulation interferes with the direct, material interests of the regulated parties is not a very good predictor of success for attempted attitude change on an issue (Sears and Funk 1991; Miller and Ratner 1998; Sanderson and Darley 2002). Note, however, that when proposed regulations would have measurable, immediate, monetary effects on the regulated parties, people *are* likely to oppose the regulation (Sears and Funk 1991).

### 3.3.3 *Effects of the Manner of Regulation*

Political scientists would argue that failures to regulate will occur, at a minimum, if advocates are insufficiently organized, interested, and funded to effectively lobby lawmakers or to withstand attacks by other organized interest groups (Olson 1965). In other words, many "attempts" at legislation that would affect moral attitudes simply never get off the ground in the earliest stages, because of a lack of coordination or resources. But even well-funded and organized attempts that succeed in changing the actual law may nevertheless fail to achieve their ultimate ambition of changing attitudes because of features of the legal regulation itself. First, if the public knows you are attempting to change moral attitudes, the regulation is less likely to work (Walster and Festinger 1962; Allyn and Festinger 1961). Effective propaganda is subtle propaganda. Relatedly, if the targets of persuasion are focused on another task, they will be more persuaded by a communication than those who are not distracted (Festinger and Maccoby 1964; Freedman and Sears 1965; Gilbert 1991). Indeed, messages that are too obvious can even induce backlashes. All of this perhaps explains the generally dismal success rates of public information campaigns (Aronson and O'Leary 1983; McGuire 1986).

While simple exhortations to change attitudes are likely to fall on deaf ears, carefully framed persuasion can work, sometimes even when it is clear to a careful observer that attitude change is the goal. For instance, people wish to conform to others they perceive to be in their in-group (Asch 1955). These effects are quite powerful—messages that are framed to appeal to self-interest or moral commitments usually lead to less behavioral (and attitude) change than those that are framed to make the listener believe that others just like them are behaving in the desired way (Cialdini 2007; Nolan et al. 2008; Goldstein, Cialdini, and Griskevicius 2008; Cialdini and Goldstein 2009). For example, rates of tax compliance go up, and deductions go down, when citizens believe *other* citizens are paying their fair share (Kahan 1997; Roth and Witte 1989).

Fellow members of a political community can be one such in-group (fellow residents of Chicago, say, or fellow Americans), and to the extent a legal regulation is seen as a reflection of democratic group consensus on an issue, it is likely to be more successful than if it were seen as a top-down imposition by elites or even outsiders (McAdams 1997; McAdams 2000a; Dharmapala and McAdams 2003). This would at least suggest that plebiscites should be more persuasive than legislation, which should be more

persuasive than judicial opinions. Research directly testing the effects of legislation or plebiscites on public opinion is rare (or nonexistent), but several studies have focused on the power of Supreme Court decisions to change attitudes (Bartels and Mutz 2009; Hoekstra 1995; Hoekstra 2000). There, the results have been mixed, and the best summation of the Court's ability to move public opinion is that it is limited, depends heavily on circumstances, and in any event is prone to inducing backlashes (Brickman and Peterson 2006; Franklin and Kosaki 1989; Johnson and Martin 1998; Persily, Citrin, and Egan 2008). Though this research is tantalizing, by referring directly to established institutions (*the Supreme Court, the US Congress*) it conflates the source-credibility issues discussed in the previous section<sup>5</sup> with the procedural/manner effects we refer to here, which have independent effects on the likelihood of failure of an official attempt to change moral attitudes.

## 4 CONCLUSION

---

In this chapter, we started from the assumption that political actors primarily wish to change behaviors, and that they often attempt to accomplish that goal—knowingly or not—by trying to change attitudes. Sometimes this works, sometimes it doesn't, and in this chapter we tried to identify when and how.

However, our initial assumption that the goal of regulation is behavioral or attitudinal change may be wrong, at least some of the time. Considerable research in political psychology suggests that at least sometimes, the point of legal regulation may instead be an attempt to co-opt the expressive capital of the law to advance a more symbolic agenda (Lau and Heldman 2009; Kinder 1998; Sears and Funk 1991). As we have written elsewhere, “Law is a form of cultural capital that can be captured by opposed groups [who use it to] stick a flag in the dirt to mark public territory as their own” (Bilz and Nadler 2009, p. 122). Social groups compete to establish laws that they perceive as expressing high social standing for their group; likewise, they oppose laws that symbolize low social standing for their group. Richard McAdams calls this view the “expressive politics theory of law” (McAdams, 2015). In this chapter, we have described a handful of regulations that could not credibly be described as anything *other* than flags in the dirt, for instance, the “nuclear-free zones” established by municipalities across the country, or the handful of affirmative “Good Samaritan” laws established in some states, which are virtually never enforced (Rosenbaum 2011, p. 248) and entail tiny penalties even when they are.<sup>6</sup>

<sup>5</sup> For instance, the Court has historically been held in high regard by the public (Caldeira and Gibson 1992), while the credibility of Congress has historically been lower (Hibbing and Theiss-Morse 1995).

<sup>6</sup> See Wisconsin Stat. Ann. § 940.34 (class C misdemeanor); Vermont Stat. Ann. tit. 12, § 519 (up to a \$100 fine); Ohio Rev. Code § 2921.22 (fourth-degree misdemeanor); Minnesota Stat. Ann. § 604A.01

Even what might look like instrumentally motivated legal regulations are often, perhaps even usually, either supported or opposed on the grounds of what they *say* rather than because of what they *do*. For instance, residents of a rural town opposed legal limits on using wood for heat to the extent they endorsed masculine pioneering ideals, despite agreeing that old-fashioned wood-burning heaters cause air pollution and health problems (Cupples, Guyatt, and Pearce 2007; Reeve et al. 2013). Support for the regulation of novel “reduced tobacco” products was predicted most strongly by the respondent’s general attitudes about the legitimate role of government in society, even when the respondent was a smoker (Kim, Stark, and Borgida 2011). Support for the Vietnam War was better predicted by “feeling thermometers” toward the military and antiwar protesters than it was by having an immediate family member serving there (Lau, Brown, and Sears 1978). The list goes on, and new data points arrive frequently.

This is not to say, of course, that self-interest is never a factor in support for legal regulations. Nor is it even to say that instrumental concerns are always swamped by symbolic ones. Indeed, as research in the area has blossomed, we have learned that the relationship among instrumental concerns, symbolic concerns, and political attitudes is quite nuanced. For example, when the actual consequences of legal regulations are immediate (Hunt et al. 2010) and clear (Sears and Funk 1991), consequences tend to dominate symbolic concerns. There are also individual differences in self-interestedness, and unsurprisingly, the more self-interested a person is in general, the more self-interest will predict her political views (Crano 1997a; Crano 1997b). The *taste* for self-interest might also be growing over time (Ratner and Miller 2001), which would suggest that in the future, the role of symbolic politics might diminish. Moreover, sometimes it’s hard to tell what is self-interest and what is “merely” symbolic: is supporting a policy that favors the group with whom you strongly identify a function of your self-interest, or of group solidarity (Luttmer 2001)?

Indeed, as the *norm* of self-interest grows—or rather, *if* it does—then this, too, should be a factor that predicts the success or failure of legal interventions. We have already noted that in Western societies built on principles of Enlightenment liberalism, people bristle at attempts by the state to impose official moral orthodoxies. Some have argued that self-interest might actually have become not just acceptable but normative as a reason to act. One study found that even altruists were reluctant to explain their own behavior as motivated by wanting to do the right thing, but instead consistently tried to describe it as motivated by instrumentalist concerns (Wuthnow 1991). If self-interest becomes the dominant grounds for why a citizen *says* she is endorsing or opposing a legal regulation, successful attempts to regulate might eventually be more likely to depend on whether they actually *do* improve a citizen’s “bottom line.”

That would be unfortunate. On the surface, in that world it would be easier to win support for a legal regulation if all advocates had to do was make it economically (in the broad sense) worthwhile to citizens. But it might be a lot more *expensive* to do so—not

(petty misdemeanor). The only exception would be Rhode Island, which assigns a fine of \$500–\$1,000 or up to six months in jail. R.I. Stat. §§ 11-1-5.1, 11-56-1.



only in the sense of “buying” support (and constructing a regulation that gives the highest material output to the largest numbers—or the most powerful numbers), but also in the sense of maintaining *compliance*. Internalized support for or opposition to a policy is stickier than support grounded in self-interest—but that stickiness is just as often an advantage as it is a disadvantage. In the grand scheme, morally motivated citizens will behave or believe as they do, *almost* no matter what the law tells or demands of them. This can be frustrating to those who wish to upset entrenched bad norms, but it can be a godsend to a legal regime with limited resources (both economic and political) to police behavior.

## REFERENCES

- Albiston, Catherine, and Shelley Correll. 2011. “Law, Norms, and the Motherhood/Caretaker Penalty.” May 30. [http://citation.allacademic.com/meta/p\\_mla\\_apa\\_research\\_citation/4/9/7/2/6/p497261\\_index.html](http://citation.allacademic.com/meta/p_mla_apa_research_citation/4/9/7/2/6/p497261_index.html).
- Alexander, Michelle. 2012. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. New York: New Press.
- Allyn, J., and L. Festinger. 1961. “The Effectiveness of Unanticipated Persuasive Communications.” *Journal of Abnormal and Social Psychology* 62 (January): 35–40.
- Aronson, Elliot, and Michael O’Leary. 1983. “The Relative Effectiveness of Models and Prompts on Energy Conservation: A Field Experiment in a Shower Room.” *Journal of Environmental Systems* 12 (3): 219–24.
- Asch, Solomon E. 1955. “Opinions and Social Pressure.” *Scientific American* 193 (November 1): 31–35. doi:10.1038/scientificamerican1155-31.
- Associated Press. 2012. “Iowa: Court Upholds Firing of Woman Whose Boss Found Her Attractive.”
- Ayres, Ian, Sophie Raseman, and Alice Shih. 2012. “Evidence from Two Large Field Experiments That Peer Comparison Feedback Can Reduce Residential Energy Usage.” *Journal of Law, Economics, and Organization* (August 20). doi:10.1093/jleo/ewso20. <http://jleo.oxfordjournals.org/content/early/2012/08/18/jleo.ewso20>.
- Barringer, Felicity. 2013. “In Kalispell, Montana, Guns Are a Matter of Life.” *New York Times*, February 20, sec. U.S. <http://www.nytimes.com/2013/02/21/us/in-montanas-kalispell-guns-are-a-matter-of-life.html>.
- Bartels, Brandon L., and Diana C. Mutz. 2009. “Explaining Processes of Institutional Opinion Leadership.” *Journal of Politics* 71 (01): 249–61. doi:10.1017/S0022381608090166.
- Becker, Gary S. 1968. “Crime and Punishment: An Economic Approach.” *Journal of Political Economy* 76: 169–217.
- Bernstein, Anita. 1996. “Better Living through Crime and Tort.” *Boston University Law Review* 76: 169–192.
- Bilz, Kenworthy, and John M. Darley. 2004. “What’s Wrong with Harmless Theories of Punishment.” *Chicago-Kent Law Review* 79: 1215–52.
- Bilz, Kenworthy, and Janice Nadler. 2009. “Law, Psychology, and Morality.” In *Moral Judgment and Decision Making*, edited by Daniel M. Bartels, Christopher W. Bauman, Linda J. Skitka, and Douglas L. Medin, 101–31. San Diego, CA: Elsevier Academic Press.



- Bohner, Gerd, and Nina Dickel. 2011. "Attitudes and Attitude Change." *Annual Review of Psychology* 62 (1): 391–417. doi:10.1146/annurev.psych.121208.131609.
- Braman, Donald, Dan M. Kahan, and David A. Hoffman. 2010. "Some Realism about Punishment Naturalism." *University of Chicago Law Review* 77: 1531–1609.
- Brawner, Dan. 2010. "City's New Nuclear Weapons-Free Zone Signs May Look Silly, but They Aren't a Joke." *Living in Iowa*. <http://mariontoday.org/times/article.php?id=1475>.
- Brehm, J. W. 1956. "Postdecision Changes in the Desirability of Alternatives." *Journal of Abnormal and Social Psychology* 52: 384–389.
- Brickman, Danette, and David Peterson. 2006. "Public Opinion Reaction to Repeated Events: Citizen Response to Multiple Supreme Court Abortion Decisions." *Political Behavior* 28 (1) (March 1): 87–112. doi:10.1007/s11109-005-9003-0.
- Caldeira, Gregory A., and James L. Gibson. 1992. "The Etiology of Public Support for the Supreme Court." *American Journal of Political Science* 36 (3) (August): 635. doi:10.2307/2111585.
- Carlson, Ann E. 2001. "Recycling Norms." *California Law Review* 89 (5) (October 1): 1231–300. doi:10.2307/3481159.
- Cassens Weiss, Debra. 2013. "Columnists Decry Iowa Ruling Finding No Sex Bias in Woman Fired for Being Too Irresistible." News. *ABA Journal*. [http://www.abajournal.com/news/article/columnists\\_decry\\_iowa\\_ruling\\_finding\\_no\\_sex\\_bias\\_in\\_woman\\_fired\\_for\\_being\\_t/](http://www.abajournal.com/news/article/columnists_decry_iowa_ruling_finding_no_sex_bias_in_woman_fired_for_being_t/).
- Cialdini, Robert B. 2007. "Descriptive Social Norms as Underappreciated Sources of Social Control." *Psychometrika* 72 (2) (June 1): 263–68. doi:10.1007/s11336-006-1560-6.
- Cialdini, Robert B. 2009. *Influence*. New York: HarperCollins.
- Cialdini, Robert, and Noah Goldstein. 2009. "Normative Influences on Consumption and Conservation Behaviors." In *Social Psychology of Consumer Behavior*, edited by Michaela Wanke, 273–96. New York: Psychology Press.
- Clarke, Alan, Jo Moran-Ellis, and Judith Slaney. 2002. "Attitudes to Date Rape and Replationship Rape: A Qualitative Study." Report of the Sentencing Advisory Panel of Great Britain.
- Clemens, Margaret A. 1982. "Elimination of the Resistance Requirement and Other Rape Law Reforms: The New York Experience." *Albany Law Review* 47: 871–907.
- Cooper, Joel, and Russell H. Fazio. 1984. "A New Look at Dissonance Theory." In *Advances in Experimental Social Psychology*, edited by Leonard Berkowitz, vol. 17:229–66. London: Academic Press. <http://www.sciencedirect.com/science/article/pii/S0065260108601215>.
- Cooter, Robert. 1998. "Expressive Law and Economics." *Journal of Legal Studies* 27: 585–608.
- Crano, W. D. 1997a. "Vested Interest, Symbolic Politics, and Attitude-Behavior Consistency." *Journal of Personality and Social Psychology* 72 (3) (March): 485–91.
- Crano, W. D. 1997b. "Vested Interest and Symbolic Politics—Observations and Recommendations: Reply to Sears (1997)." *Journal of Personality and Social Psychology* 72 (3) (March): 497–500.
- Cupples, Julie, Victoria Guyatt, and Jamie Pearce. 2007. "Put On a Jacket, You Wuss': Cultural Identities, Home Heating, and Air Pollution in Christchurch, New Zealand." *Environment and Planning A* 39 (12): 2883–2898. doi:10.1068/a3932.
- Darley, John, Tom R. Tyler, and Kenworthy Bilz. 2003. "Enacting Justice: The Interplay of Individual and Institutional Perspectives." In *The Sage Handbook of Social Psychology*, edited by Michael A. Hogg and Joel Cooper, 458–76. Thousand Oaks, CA: Sage.

- Davidai, Shai, Thomas Gilovich, and Lee D. Ross. 2012. "The Meaning of Default Options for Potential Organ Donors." *Proceedings of the National Academy of Sciences* 109 (38) (September 18): 15201–205. doi:10.1073/pnas.1211695109.
- Deboom v. Raining Rose, Inc.*, 772 N.W.2d 1, 7 (Iowa 2009).
- Dharmapala, Dhammika, and Richard H. McAdams. 2003. "The Condorcet Jury Theorem and the Expressive Function of Law: A Theory of Informative Law." *American Law and Economics Review* 5 (1) (March 1): 1–31. doi:10.1093/aler/5.1.1.
- Emmanuel, Adeshina. 2012. "The Financial Struggles of 'Nuclear-Free' Cities." *New York Times*, July 10, sec. U.S. <http://www.nytimes.com/2012/07/11/us/the-financial-struggles-of-nuclear-free-cities.html>.
- Erskine, Hazel. 1972. "The Polls: Gun Control." *Public Opinion Quarterly* 36 (3) (October 1): 455–69. doi:10.2307/2747449.
- Farley, Lin. 1980. *Sexual Shakedown*. New York: Warner Books.
- Festinger, L., and N. Maccoby. 1964. "On Resistance to Persuasive Communications." *Journal of Abnormal Psychology* 68 (April): 359–66.
- Fiedler, Paige. 2013. "Iowa View: Legislature Needs to Shut Loophole in Dentist Case." News. *Record Herald*. <http://indianolarecordherald.desmoinesregister.com/article/20130109/OPINION/301090028/Iowa-View-Legislature-needs-shut-loophole-dentist-case>.
- Foley, Ryan J. 2012. "Iowa Court: Bosses Can Fire 'Irresistible' Workers." News. *Denver Post*. [http://www.denverpost.com/commented/ci\\_22241967](http://www.denverpost.com/commented/ci_22241967).
- Franklin, Charles H., and Liane C. Kosaki. 1989. "Republican Schoolmaster: The U.S. Supreme Court, Public Opinion, and Abortion." *American Political Science Review* 83 (3) (September): 751. doi:10.2307/1962059.
- Freedman, J. L., and D. O. Sears. 1965. "Warning, Distraction, and Resistance to Influence." *Journal of Personality and Social Psychology* 95 (March): 262–66.
- Gallup. 2012. "Guns." Polling. <http://www.gallup.com/poll/1645/guns.aspx>.
- Gibson, James L., Gregory A. Caldeira, and Lester Kenyatta Spence. 2003. "Measuring Attitudes Toward the United States Supreme Court." *American Journal of Political Science* 47 (2): 354–67. doi:10.1111/1540-5907.00025.
- Gilbert, D. T. 1991. "How Mental Systems Believe." *American Psychologist* 46 (2): 107–19.
- Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius. 2008. "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels." *Journal of Consumer Research* 35 (3) (October): 472–82. doi:10.1086/586910.
- Graff, Will. 2013. "Cape Elizabeth Bans 'Filthy, Vile Habit' of Chew Tobacco for Town Employees at Work." *Bangor Daily News*, June 6. <http://bangordailynews.com/2013/03/14/news/portland/cape-elizabeth-bans-filthy-vile-habit-of-chew-tobacco-for-town-employees-at-work/>.
- Gross, Samuel, and Phoebe Ellsworth. 2001. "Second Thoughts: Americans' Views on the Death Penalty at the Turn of the Century." In *Beyond Repair? America's Death Penalty*, edited by Stephen L. Garvey, et al., 7–57. Durham, NC: Duke University Press. <http://papers.ssrn.com/abstract=264018>.
- "The Health Consequences of Involuntary Smoking." 1986. 87–8309. U.S. Department of Health and Human Services.
- Hibbing, John R., and Elizabeth Theiss-Morse. 1995. *Congress as Public Enemy: Public Attitudes toward American Political Institutions*. New York: Cambridge University Press.
- Hoekstra, Valerie J. 1995. "The Supreme Court and Opinion Change An Experimental Study of the Court's Ability to Change Opinion." *American Politics Research* 23 (1) (January 1): 109–29. doi:10.1177/1532673X9502300106.

- Hoekstra, Valerie J. 2000. "The Supreme Court and Local Public Opinion." *American Political Science Review* 94 (1) (March): 89. doi:10.2307/2586382.
- Hollander, E. P. 1958. "Conformity, Status, and Idiosyncrasy Credit." *Psychological Review* 65 (2) (March): 117–27.
- Hovland, C. I., O. J. Harvey, and M. Sherif. 1957. "Assimilation and Contrast Effects in Reactions to Communication and Attitude Change." *Journal of Abnormal Psychology* 55 (2) (September): 244–52.
- Hovland, Carl I., and Walter Weiss. 1951. "The Influence of Source Credibility on Communication Effectiveness." *Public Opinion Quarterly* 15 (4) (December 21): 635–50. doi:10.1086/266350.
- Hunt, Corrie V., Anita Kim, Eugene Borgida, and Shelly Chaiken. 2010. "Revisiting the Self-interest Versus Values Debate: The Role of Temporal Perspective." *Journal of Experimental Social Psychology* 46 (6) (November): 1155–58. doi:10.1016/j.jesp.2010.05.004.
- Jacobson, Peter D., Jeffrey Wasserman, and John R. Anderson. 1997. "Historical Overview of Tobacco Legislation and Regulation." *Journal of Social Issues* 53 (1): 75–95. doi:10.1111/j.1540-4560.1997.tb02432.x.
- Johnson, Timothy R., and Andrew D. Martin. 1998. "The Public's Conditional Response to Supreme Court Decisions." *American Political Science Review* 92 (2) (June): 299–309. doi:10.2307/2585665.
- Kahan, Dan M. 1997. "Social Influence, Social Meaning, and Deterrence." *Virginia Law Review* 83 (2) (March): 349. doi:10.2307/1073780.
- Kahan, Dan M. 2000. "Gentle Nudges vs. Hard Shoves: Solving the Sticky Norms Problem." *University of Chicago Law Review* 67 (3): 607. doi:10.2307/1600336.
- Kahan, Dan M. 2003. "The Logic of Reciprocity: Trust, Collective Action, and Law." *Michigan Law Review* 102: 71–103.
- Kahan, Dan M. 2007. "The Cognitively Illiberal State." *Stanford Law Review* 60 (1) (October 1): 115.
- Kahan, Dan M., Donald Braman, John Gastil, Paul Slovic, and C. K. Mertz. 2007. "Culture and Identity-Protective Cognition: Explaining the White-Male Effect in Risk Perception." *Journal of Empirical Legal Studies* 4 (3): 465–505. doi:10.1111/j.1740-1461.2007.00097.x.
- Katyal, Sonia K. 2006. "Semiotic Disobedience." *Washington University Law Review* 84: 489–571.
- Kauder, Neal B. 1993. "One-Gun-a-Month: Measuring Public Opinion Concerning a Gun Control Initiative." *Behavioral Sciences & the Law* 11 (4): 353–60. doi:10.1002/bsl.2370110403.
- Kim, Anita, Emily Stark, and Eugene Borgida. 2011. "Symbolic Politics and the Prediction of Attitudes toward Federal Regulation of Reduced-Exposure Tobacco Products." *Journal of Applied Social Psychology* 41 (2): 381–400. doi:10.1111/j.1559-1816.2010.00718.x.
- Kinder, Donald R. 1998. "Communication and Opinion." *Annual Review of Political Science* 1 (1): 167–97. doi:10.1146/annurev.polisci.1.1.167.
- Kinnaman, Thomas C. 2006. "Policy Watch: Examining the Justification for Residential Recycling." *Journal of Economic Perspectives* 20 (4): 219–32. doi:10.1257/089533006780387382.
- Kopel, David. 2012. "The Great Gun Control War of the 20th Century—and Its Lessons for Gun Laws Today." *Fordham Urban Law Journal* 39: 1527–616.
- Kuklinski, J. H., and Hurley, N. L. 1994. "On hearing and interpreting political messages: A cautionary tale of citizen cue-taking." *The Journal of Politics*, 56 (3): 729–51.
- Kuran, Timur. 1995. *Private Truths, Public Lies*. Cambridge, MA: Harvard University Press.

- Lacetera, Nicola, and Mario Macis. 2012. "Time for Blood: The Effect of Paid Leave Legislation on Altruistic Behavior." *Journal of Law, Economics, and Organization* (August 9). doi:10.1093/jleo/ewso19. <http://jleo.oxfordjournals.org.turing.library.northwestern.edu/content/early/2012/08/08/jleo.ewso19>.
- Langbein, Laura I. 1993. "PACs, Lobbies and Political Conflict: The Case of Gun Control." *Public Choice* 77 (3) (November 1): 551–72. doi:10.2307/30027342.
- Langbein, Laura I., and Mark A. Lotwis. 1990. "The Political Efficacy of Lobbying and Money: Gun Control in the U. S. House, 1986." *Legislative Studies Quarterly* 15 (3) (August 1): 413–40. doi:10.2307/439771.
- Lau, Richard R., Thad A. Brown, and David O. Sears. 1978. "Self-Interest and Civilians' Attitudes toward the Vietnam War." *Public Opinion Quarterly* 42 (4) (December 21): 464–82. doi:10.1086/268474.
- Lau, Richard R., and Caroline Heldman. 2009. "Self-Interest, Symbolic Attitudes, and Support for Public Policy: A Multilevel Analysis." *Political Psychology* 30 (4): 513–37. doi:10.1111/j.1467-9221.2009.00713.x.
- Lessig, Lawrence. 1995. "The Regulation of Social Meaning." *University of Chicago Law Review* 62 (3): 943–1045. doi:10.2307/1600054.
- Leventhal, Howard, Dan Meyer, and David Nerenz. 1980. "The Common Sense Representation of Illness Danger." In *Contributions to Medical Psychology*, edited by Stanley Rachman, vol. 2:7–30. New York: Pergamon Press.
- Luttmer, Erzo F. P. 2001. "Group Loyalty and the Taste for Redistribution." *Journal of Political Economy* 109 (3) (June): 500–28. doi:10.1086/321019.
- Maclay, Kathleen. 1988. "Nuclear-Free Zone Status Brings Legal, Economic Fallout to Cities." *Los Angeles Times*, February 21. [http://articles.latimes.com/1988-02-21/news/mn-44063\\_1\\_nuclear-free-zone](http://articles.latimes.com/1988-02-21/news/mn-44063_1_nuclear-free-zone).
- Maniloff, Howard. 1979. "Community Attitudes toward a Desegregated School System: A Study of Charlotte, North Carolina." Ph.D. diss., Teachers College.
- McAdams, Richard H. 1997. "The Origin, Development, and Regulation of Norms." *Michigan Law Review* 96: 338–433.
- McAdams, Richard H. 2000a. "An Attitudinal Theory of Expressive Law." *Oregon Law Review* 79: 339–90.
- McAdams, Richard H. 2000b. "Focal Point Theory of Expressive Law, A." *Virginia Law Review* 86: 1649–1729.
- McAdams, Richard H. 2015. *Expressive Powers of Law*. Cambridge, MA: Harvard University Press.
- McGuire, William J. 1986. "The Myth of Massive Media Impact: Savagings and Salvagings." *Public Communication and Behavior* 1: 173–257.
- Media Education Foundation, and FAIR (Firm). 1994. *The Date Rape Backlash: The Media and the Denial of Rape*. Videorecording. The Foundation.
- Miller, D. T., and R. K. Ratner. 1998. "The Disparity between the Actual and Assumed Power of Self-interest." *Journal of Personality and Social Psychology* 74 (1) (January): 53–62.
- Mullen, Elizabeth, and Janice Nadler. 2008. "Moral Spillovers: The Effect of Moral Violations on Deviant Behavior." *Journal of Experimental Social Psychology* 44 (5): 1239–45. doi:10.1016/j.jesp.2008.04.001.
- Mullen, Elizabeth, and Linda J. Skitka. 2006. "Exploring the Psychological Underpinnings of the Moral Mandate Effect: Motivated Reasoning, Group Differentiation, or Anger?" *Journal of Personality and Social Psychology* 90 (4): 629–43. doi:10.1037/0022-3514.90.4.629.

- Nadler, Janice. 2005. "Flouting the Law." *Texas Law Review* 83: 1399–1441.
- "Nat'l Fed'n of Indep. Bus. V. Sebelius." 2013. *Justia US Supreme Court Center*. <http://supreme.justia.com/cases/federal/us/567/11-393/>.
- Nelson V. James H. Knight DDS, P.C.* 2012, No. 11-1857. Iowa.
- Newcomb, Alyssa. 2012. "Melissa Nelson: Dental Assistant Fired for Being 'Irresistible' Is 'Devastated.'" News. *ABC News Blogs*. <http://abcnews.go.com/blogs/headlines/2012/12/melissa-nelson-dental-assistant-fired-for-being-irresistible-is-devastated/>.
- Newport, Frank. 2012. "Newtown Shootings Context: Americans' Attitudes towards Gun Control." Polling. *Gallup: Polling Matters*. <http://pollingmatters.gallup.com/2012/12/newtown-shootings-context-americans.html>.
- Nolan, Jessica M., P. Wesley Schultz, Robert B. Cialdini, Noah J. Goldstein, and Vidas Griskevicius. 2008. "Normative Social Influence Is Underdetected." *Personality and Social Psychology Bulletin* 34 (7) (July 1): 913–23. doi:10.1177/0146167208316691.
- Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.
- Opinion Research Corporation. 2011. "CNN Opinion Research Poll." <http://i2.cdn.turner.com/cnn/2011/images/06/12/new.poll.pdf>.
- Oskamp, Stuart, Maura J. Harrington, Todd C. Edwards, Deborah L. Sherwood, Shawn M. Okuda, and Deborah C. Swanson. 1991. "Factors Influencing Household Recycling Behavior." *Environment and Behavior* 23 (4) (July 1): 494–519. doi:10.1177/0013916591234005.
- Persily, Nathaniel, Jack Citrin, and Patrick J. Egan, eds. 2008. *Public Opinion and Constitutional Controversy*. New York: Oxford University Press.
- Pierce Wells, Catharine. 1996. "Date Rape and the Law: Another Feminist View." In *Date Rape: Feminism, Philosophy, and the Law*, 41–51. University Park: Pennsylvania State University Press.
- Politi, Daniel. 2012. "Iowa Supreme Court: It's Legal to Fire Employee for Being Irresistibly Sexy." News. *Slate*. [http://www.slate.com/blogs/the\\_slatest/2012/12/22/james\\_knight\\_melissa\\_nelson\\_dentist\\_fires\\_employee\\_for\\_being\\_irresistible.html](http://www.slate.com/blogs/the_slatest/2012/12/22/james_knight_melissa_nelson_dentist_fires_employee_for_being_irresistible.html).
- Ratner, R. K., and D. T. Miller. 2001. "The Norm of Self-Interest and Its Effects on Social Action." *Journal of Personality and Social Psychology* 81 (1) (July): 5–16.
- Reeve, Ian, John Scott, Donald W. Hine, and Navjot Bhullar. 2013. "'This Is Not a Burning Issue for Me': How Citizens Justify Their Use of Wood Heaters in a City with a Severe Air Pollution Problem." *Energy Policy*. doi:10.1016/j.enpol.2013.01.042. <http://www.sciencedirect.com/science/article/pii/S0301421513000682>.
- "The Respiratory Health Effects of Passive Smoking: Lung Cancer and Other Disorders." 1992. Environmental Protection Agency. [http://oaspub.epa.gov/eims/eimscomm.getfile?p\\_download\\_id=3679](http://oaspub.epa.gov/eims/eimscomm.getfile?p_download_id=3679).
- Robinson, Paul H., and John M. Darley. 1996. "Utility of Desert." *Northwestern University Law Review* 91: 453–99.
- Robinson, Paul H., and John M. Darley. 2007. "Intuitions of Justice: Implications for Criminal Law and Justice Policy." *Southern California Law Review* 81: 1–67.
- Rosenbaum, Thane. 2011. *The Myth of Moral Justice*. New York: HarperCollins.
- Roth, Jeffrey A., and Ann D. Witte. 1989. *Taxpayer Compliance*, vol. 2: *Social Science Perspectives*. Philadelphia: University of Pennsylvania Press.
- Rozin, Paul. 1999. "The Process of Moralization." *Psychological Science* 10 (3) (May 1): 218–21. doi:10.1111/1467-9280.00139.



- Rozin, Paul, and Leher Singh. 1999. "The Moralization of Cigarette Smoking in the United States." *Journal of Consumer Psychology* 8 (3): 321–37. doi:10.1207/s15327663jcp0803\_07.
- Sanderson, Catherine A., and John M. Darley. 2002. "‘I Am Moral, but You Are Deterred’: Differential Attributions about Why People Obey the Law." *Journal of Applied Social Psychology* 32 (2): 375–405. doi:10.1111/j.1559-1816.2002.tb00221.x.
- Santora, Marc. 2008. "Cigarette Tax Arrives amid Grumbling and Vows." *New York Times*, June 4, sec. New York Region. <http://www.nytimes.com/2008/06/04/nyregion/04smoke.html>.
- Schulhofer, Stephen J. 2000. *Unwanted Sex: The Culture of Intimidation and the Failure of Law*. Cambridge, MA: Harvard University Press.
- Schweder, Richard A., Nancy C. Much, Manamohan Mahapatra, and Lawrence Park. 1997. "The ‘Big Three’ of Morality (Autonomy, Community, Divinity), and the ‘Big Three’ Explanations of Suffering." In *Morality and Health*, 119–69. New York: Routledge.
- Sears, David O., and Carolyn L. Funk. 1991. "The Role of Self-Interest in Social and Political Attitudes." In *Advances in Experimental Social Psychology*, edited by Mark P. Zanna, volume 24:1–91. San Diego, CA: Academic Press. <http://www.sciencedirect.com/science/article/pii/S0065260108603275>.
- Sherif, Carolyn Wood, Muzafer Sherif, and Roger Ellis Nebergall. 1982. *Attitude and Attitude Change: The Social Judgment-Involvement Approach*. Westport, CT: Greenwood Press.
- Sherif, Muzafer, and Carl Iver Hovland. 1961. *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*. Westport, CT: Greenwood Press.
- Skitka, Linda J., and David A. Houston. 2001. "When Due Process Is of No Consequence: Moral Mandates and Presumed Defendant Guilt or Innocence." *Social Justice Research* 14 (3) (September 1): 305–26. doi:10.1023/A:1014372008257.
- Smith, Tom W. 2007. "Public Attitudes towards the Regulation of Firearms." NORC/University of Chicago.
- Sternthal, Brian, Lynn W. Phillips, and Ruby Dholakia. 1978. "The Persuasive Effect of Source Credibility: A Situational Analysis." *Public Opinion Quarterly* 42 (3) (October 1): 285–314. doi:10.2307/2748294.
- Stinchcombe, Arthur L. 1980. *Crime and Punishment—changing Attitudes in America*. San Francisco: Jossey-Bass.
- Sunstein, Cass R. 1996. "Social Norms and Social Roles." *Columbia Law Review* 96: 903–68.
- Swann v. Charlotte-Mecklenburg Board of Education* 402 U.S. 1 (1971)
- Tager, Jack. 2001. *Boston Riots: Three Centuries of Social Violence*. Boston: Northeastern University Press.
- Taslitz, Andrew E. 2005. "Willfully Blinded: On Date Rape and Self-Deception." *Harvard Journal of Law & Gender* 28: 381–446.
- Tyler, Tom R. 2006a. "Psychological Perspectives on Legitimacy and Legitimation." *Annual Review of Psychology* 57 (1): 375–400. doi:10.1146/annurev.psych.57.102904.190038.
- Tyler, Tom R. 2006b. *Why People Obey the Law*. Princeton, NJ: Princeton University Press.
- Tyler, Tom R. 2009. "Legitimacy and Criminal Justice: The Benefits of Self-Regulation." *Ohio State Journal of Criminal Law* 7: 307–59.
- Tyler, Tom R., and Jeffrey Fagan. 2008. "Legitimacy and Cooperation: Why Do People Help the Police Fight Crime in Their Communities." *Ohio State Journal of Criminal Law* 6: 231–75.
- Tyler, Tom R., and Yuen J. Huo. 2002. *Trust in the Law: Encouraging Public Cooperation with the Police and Courts*. New York: Russell Sage Foundation.

- Tyler, Tom R., and Cheryl J. Wakslak. 2004. "Profiling and Police Legitimacy: Procedural Justice, Attributions of Motive, and Acceptance of Police Authority." *Criminology* 42 (2): 253–82. doi:10.1111/j.1745-9125.2004.tb00520.x.
- Viki, G. Tendayi, Dominic Abrams, and Barbara Masser. 2004. "Evaluating Stranger and Acquaintance Rape: The Role of Benevolent Sexism in Perpetrator Blame and Recommended Sentence Length." *Law and Human Behavior* 28 (3) (June 1): 295–303. doi:10.1023/B:LAHU.0000029140.72880.69.
- Visser, Penny S, and Robert R Mirabile. 2004. "Attitudes in the Social Context: The Impact of Social Network Composition on Individual-level Attitude Strength." *Journal of Personality and Social Psychology* 87 (6) (December): 779–95. doi:10.1037/0022-3514.87.6.779.
- Walster, E., and L. Festinger. 1962. "The Effectiveness of 'Over-heard' Persuasive Communications." *Journal of Abnormal and Social Psychology* 65 (December): 395–402.
- Werner, Carol M., Jane Turner, Kristen Shipman, F. Shawn Twitchell, Becky R. Dickson, Gary V. Bruschke, and Wolfgang B. von Bismarck. 1995. "Commitment, Behavior, and Attitude Change: An Analysis of Voluntary Recycling." *Journal of Environmental Psychology* 15 (3) (September): 197–208. doi:10.1016/0272-4944(95)90003-9.
- Winkler, Adam. 2011. *Gunfight: The Battle over the Right to Bear Arms in America*. New York: W. W. Norton.
- Woods, Susan, and Douglas J. Bower. 2001. "College Students' Attitudes toward Date Rape and Date Rape Backlash: Implications for Prevention Programs." *American Journal of Health Education* 32 (4): 194–98.
- Wuthnow, Robert. 1991. "The Voluntary Sector: Legacy of the Past, Hope for the Future?" In *Between States and Markets: The Voluntary Sector in Comparative Perspective*, 3–29. Princeton, NJ: Princeton University Press
- "You're Irresistible. You're Fired." 2012. *Chicago Tribune*, December 27. [http://articles.chicagotribune.com/2012-12-27/news/ct-edit-dental-1227-jm-20121227\\_1\\_dental-assistant-dentist-discrimination](http://articles.chicagotribune.com/2012-12-27/news/ct-edit-dental-1227-jm-20121227_1_dental-assistant-dentist-discrimination).



## CHAPTER 11

---

# LAW'S LOSS AVERSION

---

EYAL ZAMIR

### 1 INTRODUCTION

---

ACCORDING to rational choice theory—the predominant theory in economics and a very influential theory in legal theory—people strive to enhance their own well-being. Among the available options, they choose the one that would maximize their expected utility, determined in absolute terms. In contrast, prospect theory posits that people generally do not perceive outcomes as final states of wealth or welfare, but rather as gains and losses (Kahneman and Tversky 1979). Gains and losses are defined relative to a baseline or reference point. The value function is normally steeper for losses than gains, indicating loss aversion. People's choices therefore crucially depend on the way they frame any choice. In particular, an individual's reference point determines whether she perceives changes as gains or losses. Usually, but not invariably, people take the status quo as the baseline. The centrality of reference points and the notion that losses loom larger than gains hold true for risky and riskless choices alike (Thaler 1980; Tversky and Kahneman 1991).

Loss aversion and related psychological phenomena, such as the endowment effect and the status quo bias, have had a significant impact on legal theory. This chapter highlights three types of contributions to legal analysis that the notion of loss aversion provides. First, inasmuch as the law strives to influence people's behavior, legal policy-makers need to understand and predict people's decision-making in legally pertinent contexts. Primary examples are consumer behavior, and litigation and settlement. The significant body of research pointing to the role loss aversion plays in consumers' and litigants' behavior is therefore of great interest to the law.

Second, various studies have established that legal norms affect the way people frame the choices they face, and in particular the reference point from which changes are perceived as losses or gains. It follows that the law may have a substantial effect on human behavior without significantly curtailing people's freedom. Two examples are contractual (and other) default rules, and the rules of burden of proof.

Third, the notions of reference points and loss aversion illuminate fundamental characteristics of the legal system itself. These notions may explain basic features of entire legal fields and even the relative importance of different fields within the legal system. Thus, for example, these notions arguably explain why, in practically all legal systems, the law of tort is far more developed than the law of unjust enrichment. Similarly, loss aversion may possibly explain the far narrower constitutional protection afforded to social and economic rights, such as education and adequate housing, compared to civil and political rights, such as the rights to life and free speech.

In addition to providing explanation and predictions of human behavior, and advancing our understanding of the legal system, each of the three interfaces between loss aversion and the law carries normative and policy implications as well. For instance, consumers' loss aversion (an instance of the first interface between loss aversion and the law) and its manipulation by suppliers may call for regulatory intervention. Likewise, if legal default rules (an example of the second interface) establish a reference point that affects people's perceptions and decisions, then the formulation of such rules may aim not only at reducing transaction costs, but possibly also at attaining redistributive and paternalistic goals. Finally, inasmuch as loss aversion impinges on people's preferences and well-being, it may not only explain basic features of the law and particular norms (the third interface), but also serve to assess their desirability as means for enhancing social welfare.

When considering these normative and policy implications, one should, however, pay heed to a fourth interface between loss aversion and the law, namely lawmakers' own loss aversion. In general, while the normative implications of loss aversion are sometimes conservative, the recognition that the law not only mirrors people's reference points, but also sometimes shapes them, may lead to progressive and even radical normative conclusions.

The chapter is structured as follows. Section 2 provides an overview of the vast psychological literature on loss aversion. Section 3 then demonstrates the contribution made by studies of loss aversion in three contexts—consumer behavior, lawyers' fees, and litigation and settlement—to legal analysis. Section 4 turns to the second interface of loss aversion and the law, framing by the law, focusing on two examples: default rules and burden of proof. Section 5 discusses in some detail the third interface: the striking correspondence between loss aversion and basic features of the law. It also briefly indicates possible explanations for this compatibility: evolutionary theories of the law and the compatibility between psychology, deontological morality, and the law. Section 6 briefly discusses some of the normative implications of loss aversion.

## 2 LOSS AVERSION: AN OVERVIEW

---

Expected utility theory assumes that people's assessment of different outcomes is independent of any reference point. Actual losses are equivalent to forgone gains (Korobkin

and Ulen 2000). This reference-independence was challenged as early as the 1950s (Markowitz 1952; Alchian 1953). It was only in the late 1970s, however, that Kahneman and Tversky (1979) came up with a full-fledged theory of people's decisions under risk—the prospect theory (PT)—that constituted a powerful alternative to expected utility theory. Thaler (1980) used PT to explain the already documented phenomenon that people attach greater value to things they have, than to things they need to acquire—which he coined the endowment effect—thus applying PT to riskless choices.

PT comprises several elements, all of which deviate from the tenets of expected utility theory. Most importantly, PT posits that people ordinarily perceive outcomes as gains and losses, rather than as final states of wealth or welfare. Gains and losses are defined relative to some reference point. The value function is normally concave for gains (implying risk aversion) and convex for losses (reflecting risk seeking). According to PT, not only the attitude to risk-taking differs in the domains of gains and losses, but the value function is also generally steeper for losses than for gains. This means that the disutility generated by a loss is greater than the utility produced by a similar gain. Since losses loom larger than gains, people are generally loss averse. The subjective value ( $sv$ ) function proposed by PT thus has the following form:

$$sv(x) = \begin{cases} x^\alpha & \text{if } x \geq 0 \\ -\lambda(-x)^\beta & \text{if } x < 0 \end{cases}$$

Under PT,  $\lambda$  (the coefficient of loss aversion) greater than 1 captures loss aversion, and parameters  $\alpha$  and  $\beta$  (assuming  $\alpha > 0$ ,  $\beta < 1$ ) capture the diminishing sensitivity to increases in the absolute payoffs. This modeling allows for the nonlinearity of the value function around the reference point. Tversky and Kahneman (1992) estimated that monetary losses loom larger than gains by a factor of 2.25. In a meta-analysis of 45 studies of the related phenomenon of endowment effect (see chapter 12 by Korobkin in this volume), Horowitz and McConnell (2002) found that the median ratio between people's willingness to pay (WTP) for a good they don't yet have, and their willingness to accept (WTA) to part with a similar good is 1:2.6 (mean 1:7.17). In a subsequent meta-analysis of 164 experiments, Sayman and Öncüler (2005) found that the median ratio between WTP and WTA is 1:2.9 (with very substantial variation).

Ordinarily, people take the status quo as the reference point, and view changes from this point as either gains or losses (Tversky and Kahneman 1991, pp. 1046–47), but this assumption is primarily appropriate for contexts in which people expect to maintain the status quo. When expectations differ from the status quo, taking people's expectations as the pertinent reference point may yield better explanations and predictions of people's behavior (Köszegi and Rabin 2006, 2007; Abeler et al. 2011). People's perception of the reference point is also influenced by the status of other people. For instance, when an employee receives a smaller salary raise than everyone else in a workplace, she may view this raise—though improving her position in absolute terms—as a loss (Kahneman and Tversky 1984; see also Loewenstein, Thompson and Bazerman 1989). Finally, one's reference point may change in dynamic situations. Most research

suggests that people quickly adapt their reference point following the making of gains (compared to their initial position), but that they are less inclined to adjust the reference point following the incurring of losses (Arkes et al. 2008).<sup>1</sup>

Loss aversion has been found in many real-world contexts and in field experiments. For example, stock rates are more volatile and riskier than treasury bills and other bonds. Yet the gap between long-term returns on stocks and on bonds is so great as to render untenable an explanation of the demand for bonds on the basis of standard notions of risk aversion (Mehra and Prescott 1985). This so-called *equity premium puzzle* is however perfectly compatible with PT's notion of loss aversion, assuming investors evaluate their portfolios annually. To avoid even a small risk of loss, people are willing to forego considerable expected gains (Benartzi and Thaler 1995). Other empirically studied patterns of behavior that may best be explained through the notion of loss aversion are taxi drivers' decisions regarding the number of hours they work each day (Camerer et al. 1997), and plaintiffs' preference for contingent fee arrangements (Zamir and Ritov 2010). Numerous other examples are readily available (Camerer 2000; DellaVigna 2009; Dittmann, Maug, and Spalt 2010). As for field experiments, a notable study was conducted by Fryer and his colleagues (2012), who tested the effect of financial incentives for teachers, based on students' achievements. While teachers in the gain group were promised a considerable monetary award for sufficient increase in their students' performance, teachers in the loss group were paid the same award in advance and asked to give back the money if their students did not improve sufficiently. Reframing failure to attain the desirable increase in students' performance as a loss, rather than as an unobtained gain, resulted in stark differences in the incentives' effectiveness. While no significant improvement was found in the gain group, the increase in the math test scores in the loss group was equivalent to increasing teacher quality by more than one standard deviation.

Loss aversion has an emotional aspect. Many studies have suggested that negative experiences have a greater impact on individuals than positive ones (Baumeister et al. 2001; Rozin and Royzman 2001). Gains and losses are closely connected to emotions of pleasure and pain (Kahneman and Tversky 1979, p. 279). The relationships between loss aversions and emotions are, however, less straightforward than one could assume. For one thing, in the context of choices among risky alternatives, it has been demonstrated that while *ex ante*, people anticipate that losses will result in a greater adverse hedonic impact than the positive impact of gains of equal magnitude, *ex post* people rationalize their losses and do not experience as great an adverse effect as predicted. It is therefore argued that, at least sometimes, the tendency to stick to the status quo results from affective forecasts rather than from affective experience (Kermer et al. 2006; Gilbert et al. 2004). Further doubts concerning the relationships between choices and emotions have been raised by experimental studies that found that predicted and

<sup>1</sup> For an alternative theory of loss aversion, focusing on the different attention people attribute to losses and gains, see Birnbaum (2008).

experienced emotional reactions to losses were not greater in magnitude than reactions to gains (Mellers et al. 1997; Mellers and Ritov 2010). However, these discrepancies plausibly resulted from the method used to elicit people's reactions (McGraw et al. 2010).

The notions of reference points and loss aversion have been used to explain such phenomena as the status quo and omission biases, the endowment effect, and sunk costs. Other things being equal, people tend to stick to the state of affairs they perceive as the status quo, rather than opting for an alternative one (Samuelson and Zeckhauser 1988; Kahneman, Knetsch, and Thaler 1991, pp. 197–99). Ordinarily, changing the status quo requires an act, while keeping the status quo involves mere omission. Hence, the tendency to keep the status quo, and the tendency to prefer omission to commission (sometimes called the default effect), are confounded. However, these biases also exist separately and their effects are additive (Schweitzer 1994). When the two biases pull in opposite directions, as when inaction is expected to result in a change, while an action is necessary to maintain the status quo, there is evidence that subjects prefer inaction (Ritov and Baron 1992).

The status quo / omission bias was experimentally demonstrated in hypothetical choice tasks (Samuelson and Zeckhauser 1988, pp. 12–21; Kahneman and Tversky 1984, p. 348), and found in surveys of the general population (Boonen, Donkers, and Schut 2011). Several natural experiments have also provided strong empirical support for this bias. Thus, Madrian and Shea (2001) studied the rate of employee participation in a retirement savings plan in a large US corporation before and after a change in the default. Before the change, employees were required to affirmatively elect participation. After the change, new employees were automatically enrolled in the plan unless they opted out of it. Changing the default resulted in a dramatic increase in retirement plan participation. Comparable data exist regarding postmortem organ donations. In some countries within the European Union, people are organ donors unless they register not to be, whereas in others they are not donors unless registered. The donation rate in most presumed-consent countries is close to 100%, while in the explicit-consent countries it varies from 4% to 27% (Johnson and Goldstein 2003).

While there are several explanations for the tendency to stick with the status quo (see, e.g., Samuelson and Zeckhauser 1988; Gal 2006; McKenzie, Liersch, and Finkelstein 2006), the primary explanation remains loss aversion (Kahneman, Knetsch, and Thaler 1991, pp. 197–99; Baron and Ritov 1994, pp. 479–80; Moshinsky and Bar-Hillel 2010; Eidelman and Crandall 2012). When departing from the status quo involves both advantages and disadvantages, people are inclined to avoid such a departure because the disadvantages likely loom larger than the advantages. For the same reason, when there is uncertainty whether departing from the status quo would result in gains or losses, people are inclined to avoid such a departure.

This explanation is closely connected to regret—the expectation that if deviation from the status quo were to result in a worse outcome than sticking with it, people would experience greater regret than if they had decided to stick with the status quo, and later learned that the alternative outcome would have been better (Kahneman and

Miller 1986, p. 145; Bar-Hillel and Neter 1996; Anderson 2003). It is also connected to the perception that people have a greater moral responsibility for harmful outcomes they actively brought about, than those they brought about passively (Sugarman 1986; Ritov and Baron 1990; DeScioli, Christner, and Kurzban 2011).

Another phenomenon associated with loss aversion is the endowment effect (sometimes called the WTA-WTP disparity). Individuals tend to place higher value on objects and entitlements they already have, compared to objects and entitlements they do not (Thaler 1980; Knetsch and Sinden 1984; Kahneman, Knetsch, and Thaler 1990). As a result, people are often reluctant to trade an item given to them for an alternative one—whatever the given item is (Knetsch 1989, 1992). The endowment effect is thoroughly discussed in chapter 12 by Korobkin in this volume.

The last phenomenon associated with loss aversion to be mentioned here is that of sunk costs and escalation of commitment. Expected utility theory posits that in choosing between different courses of action, only future costs and benefits should be taken into account. Unrecoverable, incurred costs that would not affect future costs or benefits should not affect decisions, as the past cannot be changed (Samuelson and Nordhaus 2010, p. 183). For instance, a ticket holder should decide whether to go to a concert according to the expected net benefit of doing so, regardless of how much she paid for the ticket, if at all. However, numerous laboratory and field experiments, as well as empirical studies, have established that people frequently do not disregard sunk costs. The greater the resources, time, or efforts that have already been invested in an endeavor, the more people tend to continue it (Arkes and Bloomer 1985; Staw and Ross 1989; McCarthy, Schoorman, and Cooper 1993). Various determinants—economic, organizational, social, and psychological—influence the escalation of commitment (Staw and Ross 1989; Ku 2008, pp. 222–23). Perseverance is often rational despite initial difficulties (Heath 1995); and even if it is not in the best interest of the organization, it may well be in the best interest of the decision-maker (Staw and Ross 1989, p. 216). Along with self-justification (Staw 1976; Brockner 1992; Ku 2008, p. 222), a primary psychological explanation for escalation of commitment is people's aversion to sure losses. To avoid sure losses, people tend to keep investing in failing projects even if the prospects of turning them into successful or break-even ones are slim (Thaler and Johnson 1990). In accordance with PT, sure losses are overvalued (the certainty effect), and people are risk-seeking in the domain of losses (Thaler 1980, pp. 48–49; Arkes and Bloomer 1985, pp 130–32).

Loss aversion is not necessarily eliminated by experience (e.g., Haigh and List 2005; Bokhari and Geltner 2011), nor when decisions are made by groups rather than individuals. In fact, groups may display greater endowment effect and status quo bias than individuals (Blumenthal 2012; Galin 2013).

Some studies have challenged the generality of the notions of reference-dependence and loss aversion. Evidently, different people display varying degrees of loss aversion under different circumstances. Moreover, loss aversion seems to be neutralized or even reversed for very small amounts of money (Harinck et al. 2007), and under certain experimental settings (Ert and Erev 2008). Some scholars doubt that loss aversion is the appropriate explanation for the WTA-WTP disparity (Plott and Zeiler 2005, 2007;

Gal 2006). The overall picture emerging from hundreds of theoretical, experimental, and empirical studies is clear, however. People's preferences, choices, and judgments do generally display strong loss aversion, and this phenomenon has various implications for law and legal theory.

## 3 UNDERSTANDING HUMAN BEHAVIOR IN LEGAL CONTEXTS

---

Many contexts in which people's loss aversion has been studied are legally regulated, and inasmuch as the law strives to affect people's behavior, it should take this phenomenon into account. This section uses three examples to demonstrate this claim: consumer behavior, legal fees, and litigants' behavior.

### 3.1 Consumer Behavior

Consumers' heuristics and biases may be exploited by suppliers. They are therefore of interest to legal policymakers who consider whether and how to react to such exploitation (see generally chapter 18 by Bar-Gill in this volume). Specifically, loss aversion and related phenomena may affect consumer behavior in various ways. Such an effect is expected, for instance, when prices are presented as involving discounts rather than surcharges. While economically equivalent, giving a cash discount may seem more attractive than adding a surcharge for credit, as the former frames credit as entailing a forgone gain, rather than a loss (Thaler 1980, p. 45; Hogarth 1987, p. 104; Smith and Nagle 1995, pp. 99–101). The same is true of the framing of lunch versus dinner prices and “happy hour” prices.

To trigger consumers' loss aversion and expected regret, suppliers often create the impression that if a certain good is not purchased soon, it will no longer be available or will not be available on equally favorable terms (Cialdini 2009, pp. 198–226 [marketing strategies]; see also Langevoort 1996, pp. 652–53 [investment broker techniques]). Similar effects of limited availability on purchasing behavior have been studied in the contexts of coupon expiration dates (Inman and McAlister 1994), and in-store hoarding (Byun and Sterquist 2012).

Another marketing technique that takes advantage of consumers' loss aversion is a liberal return policy. Having the right to return a product and get a refund makes the buying decision easier, since it seemingly eliminates any potential loss or regret. Should the buyer change her mind, she would be able to undo the transaction. A liberal return policy may even be seen as rendering current decision-making unnecessary, postponing it to the point in time at which the customer will decide whether to keep the product or return it. However, as Loewenstein and Adler (1995) have demonstrated,



people without an object underestimate how much they will value it once they get it. As a result of the endowment effect, buyers who own and possess the sales object value it higher; and due to the omission bias, they tend not to return it even if they are not fully satisfied with it (Hanson and Kysar 1999, pp. 733–34; Becher and Zarsky 2011). These observations may in turn raise doubts about the desirability of mandatory cooling-off periods.

A fourth role loss aversion plays in consumer transactions pertains to the documented phenomenon that consumers hardly ever read standard-form contracts (Bakos, Marotta-Wurgler, and Trossen 2009). As Becher (2007) pointed out, this reality is plausibly connected to the fact that suppliers usually present their standard-form contracts only at the end of the shopping process. By that stage, the consumer has already spent considerable time and effort in the process. According to rational choice theory, the consumer should disregard these sunk costs and decide whether to complete the transaction based on all pertinent factors, including the contract terms. However, loss-averse consumers would hate to see the time and effort they already expended as wasted, and might therefore avoid reading the contract lest they would find it unacceptable. Other heuristics and biases reinforce this tendency.

The final example mentioned here is that of contractual terms in continuous transactions that authorize suppliers to unilaterally change the transaction terms, such as updating periodic payments. Putting aside the question of why consumers acquiesce to such authorizations in the first place, one wonders why, following a detrimental change of the contract terms by the supplier, consumers who are free to terminate the contract and switch to another supplier do not often do so. Sometimes the answer is that the costs of terminating and switching are higher than their benefits. Sometimes, however, the direct costs are rather low, and it is quite clear that *ex ante* the consumer wouldn't have made the contract under the new terms. Plausibly, a central reason why consumers fail to switch to another supplier under such circumstances is their status quo and omission biases (see also Becher 2007, pp. 138–40).

Inasmuch as consumers' loss aversion, endowment effect, escalation of commitment, and the status quo and omission biases adversely affect the efficiency and fairness of consumer transactions, such inefficiency and unfairness may warrant regulation. Of course, the very existence of these heuristics and biases neither justifies any regulation nor determines its scope and content. Yet legal policymakers would be advised to consider these phenomena, rather than unrealistically assume that suppliers and consumers are all rational maximizers.

### 3.2 Lawyers' Contingent Fees

In the United States and elsewhere, contingent fee (CF) arrangements are the standard method of financing civil litigation in several types of suits, including personal injuries. Under such arrangements, the attorney's fee is contingent on the success of the claim, calculated as a certain percentage of the amount recovered, and paid upon

recovery. CF arrangements provide clients with credit (since the fee is paid only upon recovery) and assure them that they will not have to pay any fee should the claim fail. They enable plaintiffs of limited financial means to secure otherwise unaffordable legal services. The CF induces attorneys not to take cases whose expected value is too small, thus saving their clients and the courts the costs involved in pursuing such claims. It also incentivizes the attorney to win the case or obtain a favorable settlement.

The conventional flat CF rate in the United States is 33% of the recovery. The available data indicates that the effective hourly rate under CF arrangements is higher than under noncontingent (e.g., hourly) fees, and that this difference cannot be entirely attributed to the “insurance” and credit elements of contingent fees. For example, Kritzer (2002, pp. 761–68) found the mean effective hourly fee resulting from CFs to be almost twice as large as the ordinary hourly fee.

Despite their advantages, CF arrangements are hotly debated. Critics argue that the conventional rate often results in attorneys getting large fees for little work. Current CF rates, so it is charged, reflect various market failures, including plaintiffs’ inability to assess the value and prospects of their case and the scope of work involved in handling it, lawyers’ uniform pricing practices, and clients’ prohibitive search costs (Brickman 2003).

The market-failure explanations for the relatively high effective hourly fees resulting from CFs are not wholly persuasive, however. It is difficult to see how hundreds of thousands of lawyers in the United States manage to coordinate and enforce supra-competitive CF rates without any formal prohibitions on deviations from the conventional rates. An alternative economic explanation rests on the incentive role of CFs. The higher the lawyer’s share of the recovery, the more her interests converge with the client’s. To maximize their expected net recovery, clients should therefore agree to a CF rate that is higher than the one reflecting the competitive zero-profit rate, thus creating economic rents for attorneys (Santore and Viard 2001). This explanation, however, downplays lawyers’ unselfish motivations, the importance of reputation, and clients’ ability to know whether the attorney is spending the appropriate time and effort on their case (Kritzer 1990, pp. 108–11). This leads to an additional explanation for the seemingly excessive CF rates: plaintiffs’ loss aversion.

Since the outcomes of the lawyer’s work are uncertain, plaintiffs’ choice between contingent and noncontingent fees is likely conceived as a choice between two gambles. Noncontingent—hourly or global, fixed—fees expose the plaintiff to the risk of losing her case and still having to pay the attorney’s fee. It is therefore a mixed gamble in which the plaintiff may either win some gain or bear a loss. In contrast, CF is perceived as a pure positive gamble, in which the plaintiff may either gain or break even. Loss-averse people strongly prefer pure positive gambles over mixed ones, as the former do not involve the possibility of losing.

This claim assumes that plaintiffs take their position prior to hiring the lawyer as the pertinent reference point. While this is not self-evident, there are good reasons to believe that this is ordinarily the case (Rachlinski 1996, pp. 128–30; Babcock et al. 1995; Guthrie 2000, p. 182; Guthrie 2003). People internalize losses and gains and adapt

their reference points (Kahneman, Knetsch, and Thaler 1991; Tversky and Kahneman 1991, pp. 1041–42; Frederick and Lowenstein 1999). Thus, the time lapse between the point of injury or loss, and the point at which a plaintiff hires an attorney, is likely to make her view the current, post-injury/loss position as the status quo. This is not to say that plaintiffs never frame the choice of fee arrangement as belonging to the domain of losses, and such framing has indeed been manipulated experimentally (Korobkin and Guthrie 1997, pp. 95–99; Zamir and Ritov 2010, pp. 262–64). It seems, however, that these cases are rare.

Ilana Ritov and I provided experimental support for the hypothesis that plaintiffs' strong preference for CF arrangements results from their loss aversion (Zamir and Ritov 2010). We presented participants with hypothetical cases, and asked them to indicate which payment arrangement they would prefer, a CF or a fixed fee (FF). We used numerous scenarios, varying in the type of injury or loss, the expected award, the probability of winning, and the CF rate. Responders were both laypeople and experienced tort lawyers who regularly charge their clients on a CF basis (the latter were asked to state how they would act as a client seeking legal services, assuming they would not handle the case by themselves).

We found out that when the expected contingent fee (ECF)—the claimed sum multiplied by the probability of winning multiplied by the CF rate—was twice as much as the FF, 57% of the responders nevertheless preferred the CF. Even when the FF:ECF ratio was approximately 1:3.3, 43% of the subjects preferred the CF. This main effect obtained for laypeople and tort lawyers, and across a broad range of fees. Loss aversion—rather than risk aversion—explains why presumably richer lawyers displayed the same preferences as poorer students, and why the preference for CF was manifest both when the expected fees were quite high and very low. The preference for CF was not eliminated when the incentive effects of the different fee arrangements were neutralized (Zamir and Ritov 2010, pp. 264–67).

Our findings do not refute the claim that the prevalence of CFs and their high rates are due to clients' information problems; yet they considerably weaken this claim for two reasons. First, in all the experiments, subjects were provided with all necessary information and were nevertheless willing to pay much higher expected fees under the CF arrangement than under the noncontingent arrangement. Second, the choices made by experienced tort lawyers who are very knowledgeable about CFs were practically identical to those of laypeople.

These findings imply that CF arrangements provide loss-averse plaintiffs with a significant advantage. The findings cast doubt on the idea that imposing disclosure duties on lawyers would make a substantial difference in market behavior. The same holds true for requiring lawyers to offer their prospective clients a choice between different fee arrangements. Despite the explicit availability of such alternative options in our experiments, respondents manifested a strong preference for CFs. The fact that the preference for CF also characterizes people who face no cash constraints is a relevant factor in considering the distributive effects of CF and its regulation. Interestingly, while consumers' loss aversion and related phenomena may call for regulation in

consumer markets, in the context of CF clients' loss aversion arguably weighs against regulation.

### 3.3 Litigation and Settlement

Ever since the 1970s, a vast number of economic studies have analyzed litigation behavior based on the assumption that litigants are rational maximizers, who decide whether to file suit, litigate, or settle according to the expected gross return from litigation and its expected costs (for a review, see Spier 2007). More recently, behavioral studies have shed new light on this sphere (see generally chapter 24 by Robbennolt in this volume). One line of investigation has highlighted the obstacles for settlement even when settlement would be mutually beneficial (Korobkin and Guthrie 1994; Babcock et al. 1995; Rachlinski 1996; Guthrie, Rachlinski, and Wistrich 2001). These studies have shown how people's (primarily defendants') inclination to take risks in order to avoid sure losses may hinder mutually beneficial settlements. However, the combination of regret and loss aversion provides a more direct and more powerful explanation for the fact that the great majority of cases *do* settle.

People experience an unpleasant feeling of regret when, following a decision they make, they realize that they could have obtained a better outcome had they made a different decision. Anticipation of possible ex post regret affects people decisions ex ante (Bell 1982; Loomes and Sugden 1982; Larrick and Boles 1995). Other things being equal, people prefer an option that will not expose them to an ex post feeling of regret, or will expose them to such a feeling to a lesser extent.

In and of themselves, neither the regret theories proposed by Bell (1982) and by Loomes and Sugden (1982), nor the more recent decision affect theory proposed by Barbara Mellers and her coauthors (Mellers et al. 1997; Mellers, Schwartz, and Ritov 1999), hinge on the decision-maker's loss aversion. However, loss aversion intensifies the effect of anticipated regret. Whenever the forgone option may turn out better or worse than the chosen one, the chooser may anticipate either ex post regret, or ex post rejoicing. Loss aversion implies that the potential regret would loom larger than the potential rejoicing (Larrick and Boles 1995, p. 89). Anticipated rejoicing is thus unlikely to offset anticipated regret (Larrick and Boles 1995, p. 89; Mellers, Schwartz, and Ritov 1999).

Facing a choice between two (or more) options, the anticipation of regret primarily depends on what the decision-maker expects to know ex post. She may know the outcomes of neither the chosen option nor the unchosen one (no knowledge), the outcome of both options (full knowledge), or the outcome of the chosen option only (partial knowledge) (Ritov and Baron 1995). Under the circumstances of full knowledge, the outcome of the forgone option is a salient reference point; and if it turns out to be better than the outcome of the chosen option, one is likely to feel regret. Regret is not inconceivable with partial knowledge, as the decision-maker may imagine what the

outcomes of the unchosen alternative might have been; yet it is clearly less likely in that case (Mellers, Schwartz, and Ritov 1999).

The options of litigation and settlement are asymmetric in terms of the decision-maker's expected knowledge. The litigant either accepts a settlement offer, a certain result, or goes to trial—whose results will never be known if the parties settle the case. *Ceteris paribus*, under such circumstances a loss- and regret-averse person will likely find settlement considerably more attractive than litigation, since not knowing the outcome of litigation shields one from the anticipated regret. Regret- and loss-aversion thus explain why so many litigants (plaintiffs and defendants alike) prefer to settle their cases despite the prevalence of asymmetric information, strategic considerations, and a host of psychological barriers to settlement (Guthrie 1999). Guthrie (1999) provided some experimental and anecdotal support for this claim, as well as advice to lawyers whose clients or adversaries are regret- and loss-averse.

## 4 FRAMING BY THE LAW

---

Whereas the previous section argued that people's loss aversion is a factor policymakers should take into account in designing legal norms (loss aversion as an input, if you like), the present section discusses cases in which the law affects human behavior by creating and using loss aversion (loss aversion as an output). Two examples for this interface of law and loss aversion are mentioned: default rules and the burden of proof in litigation.

### 4.1 Default Rules

According to standard economic analysis, default rules primarily serve to save on transaction costs. Thus, if most parties to a certain type of contract ordinarily prefer a certain arrangement, the law may save them the cost of negotiating and incorporating such an arrangement into their contract by laying down that arrangement as a default rule. Default rules constitute a large part of contract law, inheritance law (determining how the estate will be divided in the absence of a will), family law (determining the spouses' rights in the absence of a prenuptial agreement), and health law (determining whether people will be postmortem organ donors in the absence of explicit consent or refusal to be ones). Standard economic explanations for why people are reluctant to deviate from the status quo, even when such deviation might be beneficial, focus on factors like transaction costs, the concern that courts would err in interpreting idiosyncratic terms (Goetz and Scott 1985), and the undesirable signaling liable to result from initiation of such deviation (Ben-Shahar and Pottow 2006). However, as discussed in greater detail in chapter 12 by Korobkin in this volume, the status quo bias seems to play a central role in this context as well. Legal norms set the pertinent baseline for

people who are subject to those norms, thereby affecting the behavior of loss-averse people (Abadie and Gay 2006; Rachlinski and Jourden 1998; Korobkin 1998).

## 4.2 Burden of Persuasion in Civil Litigation

The burden of persuasion in civil litigation determines which party prevails when all the evidence has been introduced and the trier of fact is left in doubt. In such a case the fact-finder should decide against the party who bears the burden of persuasion. The basic rule in civil litigation in the Anglo-American tradition is “preponderance of the evidence” or “balance of probabilities,” conventionally meaning that the plaintiff has to persuade the fact-finder that the existence of the facts giving rise to her alleged legal right are more probable than their nonexistence (e.g., McCormick 2006, p. 484). In some specific contexts, such as allegations of fraud, US law sets a higher standard of persuasion: “clear and convincing evidence” (McCormick 2006, pp. 487–90). In addition, various legal presumptions, resting on a range of substantive and procedural considerations, place the burden of persuasion regarding particular issues on the defendant or set specific standards of persuasion for particular facts (McCormick 2006, pp. 495–522).

According to common wisdom, the only function that burden of persuasion plays in civil cases is to resolve ties (Friedman 1997, p. 1970; Stein 1998, p. 319; Johnson 1999, pp. 438–39, 444–46). If this were true, the burden of persuasion would only be significant in cases of insufficient evidence, where the incomplete evidentiary basis (or the total lack of evidence) would thwart any factual determination. Whenever the parties present a sufficient amount of evidence, burden of persuasion would play a role “only when the evidence is in perfect equipoise, in other words (since perfect equipoise is chimerical), virtually never” (Johnson 1999, p. 438).

Why, then, does the burden of persuasion attract the great doctrinal and theoretical interest that it does, and why do legislators and courts devote considerable time and energy to promulgating various rules placing and shifting it from one party to another? One answer is that the burden of persuasion is often associated with the burden of producing evidence, and the latter does significantly affect outcomes whenever neither party has access to good evidence (Allen 1994, pp. 639–40). A more direct and at least equally important answer is that evidentiary presumptions and the burden of persuasion are not mere tiebreakers. Rather, they serve as reference points.

There is no reason to assume that the psychology of judges and other fact-finders is fundamentally different than that of other people (Prentice and Koehler 2003, pp. 638–39; Guthrie and George 2005, pp. 374–85), and some studies have indeed demonstrated that it is not (Guthrie, Rachlinski, and Wistrich 2001; see generally chapter 26 by Teichman and Zamir in this volume). It thus stands to reason that legal norms impact judicial decisions not only through their direct and explicit content, but also through their more subtle effect on the way judges and juries frame the pertinent issues.

The rules regarding burden of persuasion affect fact-finding even when there is sufficient evidence by establishing a benchmark or a reference point. As Karl Llewellyn (1937, p. 385 n. 114) insightfully observed more than seventy-five years ago, “burden of proof is more than burden of proof; it also indicates base-lines for judging proof, when offered.” Unless the burden is discharged, the legal default assumption is that the alleged fact does not exist. The party bearing the burden of persuasion thus has to overcome the omission bias and the default effect created by the rule.

This understanding of the burden of persuasion is compatible with experimental findings concerning the actual meaning of the preponderance-of-the-evidence rule. In a series of experiments, Ilana Ritov and I presented legally trained subjects with various descriptions of legal disputes and the evidence produced by the parties. Half of the subjects were asked to rate the persuasiveness of the plaintiff’s version on a scale of 0 to 100, where 0 indicates that there is no doubt that the plaintiff’s version is incorrect and 100 indicates that there is no doubt that it is correct (the scale question). The other half were asked to indicate whether they would accept the claim or reject it (the decision question). We found that the percentage of responders in the decision version who accepted the claim was considerably smaller than the percentage of respondents in the scale version who rated the persuasiveness of the plaintiff’s version as above 50. We also computed the cutoff point on the scale, above which the percentage of responders in the scale condition corresponded to the percentage of responders accepting the claim in the decision condition (for example, if 31% of the responders in the decision question accepted the claim, we calculated the minimal rating of the top 31% of the responders in the scale question). The cutoff point hovered around 70 (Zamir and Ritov 2012, pp. 176, 177, 180, 186–87). Our experiments lent support to the conjecture that the gap between the formal meaning and the actual impact of the preponderance-of-the-evidence rule is due to fact-finders’ omission bias (Zamir and Ritov 2012, pp. 180–82, 192–93).

As legal default rules and the burden of persuasion demonstrate, people’s loss aversion is not only an aspect of human decision-making the law should take into account; it is also a phenomenon the law can use to attain various results. This possibility carries normative implications that are discussed in section 6. Before turning to this discussion, however, the following section discusses the intriguing compatibility between loss aversion and basic features of the legal system itself.

## 5 LOSS AVERSION AND BASIC FEATURES OF THE LAW

---

Drawing on Zamir (2012), this section argues that the law not only takes people’s loss aversion into account and affects people’s behavior by framing their choices (as argued in sections 3 and 4, respectively); the law itself is also compatible with the notion of loss aversion. The section also discusses possible explanations for this compatibility.



## 5.1 Compatibility between Loss Aversion and the Law

This subsection uses two examples—the dissimilar role that tort law and the law of unjust enrichment play in all legal systems and the greater protection afforded to civil and political human rights compared to social and economic human rights—to demonstrate the compatibility between loss aversion and the law. Additional illustrations will be briefly mentioned.

### 5.1.1 *Private Law: Tort versus Unjust Enrichment*

A common feature of practically all legal systems is the manifest gap between the centrality of the law of tort and the relative marginality of the law of restitution and unjust enrichment (Gordon 1992, p. 450; Epstein 1994, pp. 1369–71). Interactions in which one person suffers injury or loss due to another person's conduct give rise to legal entitlements and remedies far more frequently than interactions in which one person receives a considerable benefit thanks to another person's conduct (Levmore 1985). Relatedly, when the same interaction results in both injury to one side and benefit to the other, the injured party's remedial rights are usually based on her losses, rather than on the other party's gains (Grosskopf 2001, pp. 1994–95).

However positive and negative externalities are defined, from an economic perspective the actor should internalize both types of externalities to induce efficient behavior. Yet the law of tort requires injurers to pay for their negative externalities much more often than the law of unjust enrichment entitles benefactors to regain the benefits they confer upon others. Various explanations have been offered for this puzzle. The economic explanations have focused on such factors as the alleged differences between involuntary injuries and benefits in terms of the feasibility of bargaining between the parties (Levmore 1985, pp. 79–82), and the expected result of affording veto power to people who are affected by the relevant (injurious or beneficial) activity (Porat 2009). Other explanations refer to the valuation difficulties courts would face were a broad right for the restitution of benefits recognized (Levmore 1985, pp. 69–72; Gordon 1992, pp. 456–57).

Without necessarily rejecting these explanations, loss aversion seems to provide a particularly strong explanation for the current puzzle. If losses loom larger than gains, and if the parties' positions prior to the infliction of the loss or the bestowing of the benefit are the natural reference point, then a person who suffered a loss is much more likely to seek redress from the injurer than a person whose behavior yielded an unrequested benefit for some beneficiary. While the former redress is likely viewed as a remedy for a loss, the latter is more likely viewed as trying to obtain a gain. From the viewpoint of a disinterested arbiter, such as a judge or a legislator, compensating the injured person for her strongly felt loss is seen as much more pressing than entitling the provider of the nonbargained benefit to recover for the less-strongly-felt unattained benefit.

### 5.1.2 *Human Rights: Civil and Political versus Social and Economic*

Human rights are fundamental rights to which all human beings are entitled qua human beings (Donnelly 2003, pp. 7–21). Within this broad category, a basic distinction

is drawn between civil and political rights (CPRs), and social and economic rights (SERs). CPRs include rights and liberties such as the rights to life, bodily integrity, freedom of speech and religion, and the right to participate in the political process. SERs include the right to an adequate standard of living, including adequate nutrition, clothing, and housing. They also include the rights to medical services and education, the right to work, and the right to property.

In many jurisdictions the scope of constitutional protection afforded to SERs is far narrower than the constitutional protection of CPRs, if it exists at all (Sunstein 2005; Gardbaum 2008, pp. 446–53; Gavison 2003). Various explanations have been offered for the distinction. It is most commonly associated with the distinction between negative and positive rights. CPRs are perceived as merely requiring the state to refrain from certain acts, whereas SERs are considered as entailing positive duties and substantial public expenditure (Cranston 1973; Cross 2001). Deontological morality prioritizes the prohibition on actively/intentionally harming other people over the duty to promote human welfare (cf. Zamir and Medina 2010, pp. 41–48, 57–78). Contrary to this argument, however, protecting CPRs often requires positive steps and considerable costs (Holmes and Sunstein 1999; Gavison 2003, pp. 33–35; Gardbaum 2008, pp. 444–46, 453–61). For example, to guarantee the freedom of assembly, the police may have to allocate large resources to protecting demonstrators from being attacked by their opponents.

Institutionally, effective protection of human rights usually entails judicial review of legislation and administrative actions. Courts arguably possess the professional competence and enjoy the legitimacy necessary to identify and prevent CPR violations, but they lack the macroeconomic data, skills, and legitimacy needed to define the scope of SERs and enforce them (Cross 2001). In response, it is argued that judicial enforceability is not a precondition for the recognition of human rights, and that delineating the scope of human rights inescapably entails value judgments and impinges on the allocation of public resources, whether it pertains to CPRs or to SERs (Gavison 2003).

A powerful alternative or complementary explanation for the lesser protection of SERs—though not necessarily a justification—rests on the gains/losses distinction. Both when the state refrains from silencing people or taking their land, and when it takes positive measures to protect free speech against suppression by other people or to protect private property from intruders, it prevents a loss or harm to the speaker or landowner. At the same time, freedom of speech does not necessarily require the government to provide therapy for people with speech impediments or to facilitate access to communications media. In the sphere of SER, the provision of housing or health services is far more likely to be perceived as giving people something they do not have, and thus as belonging to the domain of gains.

This conjecture is compatible with cases in which legal systems that do not generally protect social rights nevertheless impose positive duties on the government. Thus, in *Goldberg v. Kelly* (397 U.S. 254 (1970)) the Supreme Court of the United States held that due process forbids the *termination* of welfare benefits without fair hearing. In a similar vein, the notion of reference point may also help explaining why it is that—while the Constitution does not mandate the provision of economic benefits—once such benefits

are provided to some segments of society, they must be provided without discrimination to similarly situated people (Currie 1986, pp. 881–82). The fact that certain benefits are granted to some people may change the reference point of similarly situated people, such that those who do not receive them experience this as a loss.

### 5.1.3 *Additional Examples*

The compatibility of law and loss aversion manifests itself in a host of additional spheres, of which I will mention but a few. One such sphere is remedies for breach of contract. While the law readily compensates the injured party for its losses (through either expectation or reliance damages), it does not ordinarily entitle her to the breacher's gains from the breach (the so-called disgorgement remedy) (Zamir 2012, pp. 852–60). Another sphere is affirmative action. Affirmative plans are highly controversial. However, there seems to be a consensus that, even if they are justifiable in hiring procedures (commonly perceived as involving a gain), they can hardly ever be justified in firing procedures (perceived as inflicting a loss) (Zamir 2012, pp. 860–64). Moving from private law and civil rights to criminal law, conduct that the actor believes to be necessary to avoid harm or evil is justifiable under certain circumstances, but no such justification applies to conduct believed to be necessary to produce benefit or good (Zamir 2012, p. 868). Under international and domestic refugee law, once asylum seekers are physically present within a country, they enjoy various substantive and procedural rights. It is much more controversial, however, whether and to what extent countries may legitimately prevent asylum seekers from ever reaching their territory. Expelling physically present people is likely perceived as inflicting a loss, while denying a visa and other pre-entry devices is seen as not providing a benefit (Zamir 2012, p. 866).

Having presented a few examples of the compatibility between loss aversion and the law, the following subsections offer two possible explanations for this compatibility.

## 5.2 Law and Psychology: Evolutionary Theories

Legal economists have long argued that by and large the common law is efficient; and one of the explanations for this observation has been evolutionary. Starting with the seminal articles of Rubin (1977) and Priest (1977), an extensive body of literature has examined the hypothesis that even if judges do not care about efficiency, a process in which inefficient rules are gradually extinguished while efficient ones survive may result from the self-serving behavior of litigants (see generally Rubin 2005; Parisi 2004). While all versions of this hypothesis have been sharply criticized, this body of literature contains valuable insights that may shed light on the compatibility between loss aversion and basic features of the law. One such insight is that the direction of the law's evolution is not only set by courts' reasoned decisions; the behavior of the litigants is important as well (Priest 1977; Goodman 1978). Another insight is that, even prior to the decision of whether to litigate or settle a case, the existence of a dispute is a

precondition for the evolution of judge-made law, efficient or not (Stake 1991, p. 1492). No judge-made rule could evolve absent a legal dispute.

The economic literature assumes that whenever there is a legal dispute, the parties decide whether to litigate or settle out of court according to the expected costs and benefits of each alternative. If, however, people perceive losses as much more painful than unobtained gains, then potential plaintiffs would be much less inclined to sue for unobtained gains than for losses. Since unobtained gains are less likely to produce disutility large enough to justify legal action (which typically entails high costs—direct and indirect, pecuniary and nonpecuniary), considerably fewer disputes are expected to revolve around unobtained gains. Since legal norms develop out of disputes, it stands to reason that the law of unjust enrichment and disgorgement remedies, to name but two examples, would be considerably less developed than the law of torts and reliance or expectation remedies. While this hypothesis focuses on judge-made law evolving as a result of the behavior of plaintiffs who are single-shot players, it basically holds true when we turn our attention to the “supply side” of precedents (competition between courts), consider plaintiffs who are repeat players (but who may nevertheless be loss averse), and reflect on statutory law (that is also affected by the demand for legal norms) (Zamir 2012, pp. 873–75).

The evolutionary theory is subject to criticism. First, even if losers are more likely to file suits than no-gainers, one should only expect significant difference in the rate of lawsuits with regard to relatively small gains. Hence, the evolutionary theory may explain why legal norms dealing with losses develop faster than norms dealing with unobtained gains, but it does not necessarily account for the dramatic asymmetries observed in subsection 5.1. At the same time, if one takes the analogy between legal and biological evolution seriously, and given that the resources of litigants, lobbyists, and legal policymakers are limited, then one would expect that the greater resources devoted to developing doctrines protecting people from losses would crowd out doctrines dealing with unobtained gains. Also, while norms dealing with losses are more developed than those dealing with unattained gains, the latter do exist.

The evolutionary theory may also be criticized on the grounds that from the plaintiff's perspective, legal relief may always be perceived as belonging to the domain of gains (see subsection 3.2). It nevertheless remains true that people who incurred a loss are more strongly motivated to seek legal redress than people who failed to obtain a gain.

These and other criticisms call for cautiousness and modesty. The evolutionary theory of the compatibility between law and loss aversion does not purport to explain the intricacies of any legal field. Rather, it is offered as an explanation for general, basic features of the law. It is modest also in the sense that it is not offered as the exclusive or even a primary explanation for the compatibility between loss aversion and the law. It is only meant to be supplementary to an explanation focusing on the mindset of legal policymakers, to which I now turn.

### 5.3 Cognitive Psychology, Commonsense Morality, and the Law

While the evolutionary hypothesis is plausible, a more robust explanation for the compatibility between the psychological notion of loss aversion and the law seems to rest on an intermediate factor: commonsense morality. This explanation posits that by and large, the law conforms to prevailing moral intuitions, and since the latter are closely linked to notions of reference points and loss aversion, these notions shape the law as well.

Briefly, commonsense morality is deontological. People believe that enhancing good outcomes is desirable, yet they also hold that attaining this goal is subject to moral constraints. These constraints include prohibitions against lying, breaking promises, and most importantly, against intentionally/actively harming other people. It is immoral, for example, to kill one person and harvest her organs to save the lives of three other people, even though the benefit of such killing (saving three people) outweighs the cost (killing one person) (Kagan 1998, pp. 70–78; Zamir and Medina 2010, pp. 41–42). The compatibility between commonsense morality and (moderate) deontology has been pointed out by deontologists and consequentialists alike, such as Scheffler (1988, p. 9), Brennan (1995, p. 145), and Kagan (1989, pp. 1–5).

Deontological morality distinguishes between harming a person and not benefiting her. Were promoting the good as compelling as eliminating the bad, the doing/allowing and the intending/foreseeing distinctions, which are essential for the deontological moral constraint against harming people (or at least one of the two is), would have collapsed. According to these distinctions, while it is forbidden to intentionally/actively harm people, there is a considerably less stringent constraint against merely foreseeing or allowing people to suffer an injury or a loss (see generally Kagan 1998, pp. 94–100; Kagan 1989, pp. 83–127; for psychological studies substantiating the prevalence of this intuition, see Ritov and Baron 1990; Baron and Ritov 1994). The prohibition against killing one person in order to save the lives of three other people necessarily implies that intentionally/actively killing an involuntary “donor” is worse than merely foreseeing/allowing the death of three people. Otherwise, there would be a prohibition against both killing the one and not killing her (thus foreseeing/allowing the death of the three). Now, whenever an agent abides by the prohibition against intentionally/actively doing harm (e.g., she refrains from killing one person), she simultaneously avoids *intending/doing harm* to the one and avoids *intending/doing good* to the three. The intending/foreseeing and doing/allowing distinctions thus inevitably entail an intending good/intending bad and doing good/doing bad distinctions. Promoting the good is less morally compelling than eliminating the bad (on this distinction, see Kagan 1989, pp. 121–25 [a critique]; Kamm 1992, pp. 381–82 [a defense]).

The moral distinction between promoting the good and eliminating the bad corresponds straightforwardly with the psychological notions of reference points and

loss aversion. Losses, unhappiness, disutility, and harm loom larger than gains, happiness, utility, and benefit.<sup>2</sup> Indeed, most of the psychological studies have focused on people's perceptions and choices regarding gains and losses to themselves, whereas morality centers on the effects of one's conduct on other people. However, as Ritov and Baron (1990, pp. 483–89), Vars (1996), and Moshinsky and Bar-Hillel (2010), among others, have demonstrated, loss aversion characterizes not only people's perceptions and choices regarding their own health, wealth, or welfare, but also regarding the effects of one's decisions on the health, wealth, or welfare of others. Therefore, even if one sets aside the evolutionary explanations based on plaintiffs' behavior, the prevailing moral intuitions of legal policymakers—legislators, judges, and administrators—may explain the manifest correlation between psychology and law described in subsection 5.1.

In addition to the close correspondence between psychology and morality, this thesis assumes a correlation between morality and law. There is indeed a broad consensus, cutting across different theories of law, that such correlation does exist (Zamir 2012, pp. 880–81). In fact, the basic features of the law discussed in subsection 5.1 correspond to the distinction between doing bad (inflicting a loss) and doing good (conferring a gain), presupposed by the deontological doing/allowing and intending/foreseeing distinctions.

The proposed correspondence between psychology, morality, and law falls into line with recent theories of evolutionary psychology, evolutionary morality, and moral psychology, which exceed the scope of the present discussion (see Zamir 2012, pp. 881–84). Inasmuch as it can be established that loss aversion is a universal characteristic of human psychology, and that the basic features of the law described in subsection 5.1 universally characterize all legal systems—or at least those systems that are sufficiently advanced to deal with the pertinent issues—these universalities arguably lend support to theories of moral psychology. My argument does not, however, hinge on these theories.

To sum up, the correlation between the psychological phenomena of reference-dependence and loss aversion, and basic characteristics of the law are not coincidental. This correlation may be the product of the evolution of judicial and statutory law, given plaintiffs' (and interest groups') stronger motivation to seek redress for losses than for unobtained gains. Primarily, though, it reflects the mindset of legal policymakers, whose moral intuitions conform to commonsense morality. Commonsense morality treats harms and benefits very differently. Just as psychologically, losses loom larger than gains, normatively, harms loom larger than benefits.

<sup>2</sup> One may also observe a correspondence between the moral doing/allowing distinction (which is closely connected to the doing good / doing bad distinction) and the psychological omission bias (closely connected to status quo bias and loss aversion). On omission bias and loss aversion see Bar-Hillel and Neter (1996, p. 25).



## 6 NORMATIVE IMPLICATIONS

---

Each of the three interfaces between the psychological phenomenon of loss aversion and the law carries normative implications. Theories of human psychology and their interaction with the law may contribute to normative deliberation in at least two ways. First, human psychology is relevant for the construction of a normative theory. Basic elements of any normative theory, including its underlying theory of human well-being, are founded on assumptions about human psychology. Second, once a normative theory is formulated, legal policymakers aiming at a certain goal, such as the promotion of economic equality or deterrence of antisocial behavior, face pragmatic choices between different means of achieving that goal. Positive theories of human psychology prove essential in making these choices. I will look at these two issues in turn, and then address possible reservations about the normative implications of loss aversion.

### 6.1 Loss Aversion and Human Welfare

All normative theories take outcomes into account, whether as the only factor that ultimately determines the morality of an act, rule, or anything else (consequentialism), or as one of several such factors (deontology). The one type of outcome all theories deem relevant is the effect of any act, rule, or anything else on human welfare (Kagan 1998, p. 70). Theories of human welfare may be grouped under three large categories: hedonistic, preference-based, and objective list (Parfit 1984, pp. 493–502; Kagan 1998, pp. 25–41). Whichever theory of human welfare one adopts, inasmuch as loss aversion affects people's mental state and their preferences, it must be taken into account. Even under objective-list theories of human well-being, any plausible list of objective goods includes enjoyment (experiencing pleasure and avoiding pain) and personal freedom (the ability to set one's goals and fulfill one's desires) as important features of well-being (Lewinsohn-Zamir 2003, pp. 1703–5). Hence, although they may assign lesser weight to loss aversion (compared to mental-state, hedonistic theories and preference-satisfaction theories), objective-list theories should also take loss aversion into account.

Arguably, this conclusion may not hold under ideal-preference theories, which posit that only the satisfaction of ideal (rather than actual) preferences enhances people's well-being. This claim assumes that loss aversion is an irrational bias. Thus, even if it typifies people's actual perceptions and preferences, an ideal-preference theory should disregard preferences stemming from loss aversion. This claim further assumes that ideal preferences are rational, and presupposes a particular notion of rationality, namely, expected utility maximization. These assumptions are, however, problematic. Nothing in expected utility theory necessitates a reference-independent



utility function. Just as a utility function may reflect risk aversion, risk neutrality, or risk seeking (Savage 1972, pp. 103–4), it may reflect either reference-independence or reference-dependence. To be sure, some related aspects of decision-making, such as its susceptibility to framing effects and an extreme, paralyzing aversion to losses, may indeed be deemed irrational. But as dozens of studies have shown, errors of all sorts characterize decision-making, independent of whether one's utility function is reference-dependent. Reference-dependence and loss aversion are not irrational *per se*. If reference effects shape the experienced value of outcomes, then it is perfectly rational to take those effects into account (Fischer et al. 1986, pp. 1082–83; see also Rubinstein 2006, pp. 107–11). Money and other resources are instrumental for attaining intrinsic goods. If the carriers of utility are not final levels of wealth, but rather gains and losses, then ideal-preference theories need not disregard loss aversion.

## 6.2 Loss Aversion as an Input to, and a Means of, Legal Policymaking

Section 3 showed that human behavior in legally important spheres, such as consumer transactions and litigation, can be better understood and predicted once loss aversion is taken into account. Section 4 demonstrated that to better understand the effect of legal norms, such as default rules and the burden of persuasion, one has to take into account the framing effect of these norms. This does not mean that one can draw any direct or conclusive normative conclusion from loss aversion. For one thing, the scope of loss aversion and its interaction with other heuristics in any particular context vary considerably. For another, the normative picture is always complicated by a myriad of competing and conflicting normative arguments, institutional considerations, and pragmatic constraints. Nevertheless, once loss aversion is established, it cannot be ignored. This is true of market transactions, where the possible manipulation of consumer's loss aversion and related phenomena by suppliers may justify regulation; as it is true of the market for legal services, where client's loss aversion may militate against regulation of contingent fees. Loss aversion and the status quo / omission bias also open the door to very mild forms of regulation that attain far-reaching results (Thaler and Sunstein 2009; and see below).

## 6.3 Loss Aversion as a Justification for Basic Features of the Law

The compatibility of law and loss aversion described in section 5 has a normative aspect as well. Loss aversion not only explains basic features of the law and particular legal norms, it can also justify these features. The greater efforts the law puts into deterring the infliction of losses and remedying losses once incurred—compared to incentivizing the conferring of benefits and rectifying the nonattainment of gains—make perfect sense as a means to enhance human welfare, since losses typically loom larger than

gains. This observation also implies that loss aversion may serve to assess the desirability of existing and proposed new legal norms.

## 6.4 Normative Implications of the Manipulability of Reference Points

The ability of the law to set reference points, as discussed in section 4, arguably diminishes the inclusion of people's loss aversion as an input to legal policymaking (discussed in section 3) and as an explanation and justification for extant legal norms (described in section 5 and subsection 6.3). If the law can effortlessly and effectively shape people's reference points, then preexisting baselines could just as well be ignored.

This argument, however, proves too much. While people's framing is sometimes malleable, the ability of the law to frame people's perceptions varies from one context to another and is often limited. A pertinent example is tax deductions and credits. Following a powerful campaign pointing to the various pitfalls of using these measures, the federal government and most states in the United States have adopted "tax expenditure budgets" that identify and quantify tax subsidies, credits, and deductions as expenditures (Surrey and McDaniel 1985, pp. 1–30, 238–39; Zelinsky 2005, p. 803). It was expected that once the true nature and problematic ramifications of tax subsidies became apparent, legislators would refrain from channeling expenditures through the tax system. This assumption was in the main proven wrong, as the use of tax expenditures has not been reduced (Thuronyi 1988, pp. 1170–81; Zelinsky 2005, pp. 801–4; Kleinbard 2010). Apparently, legislators and the public at large keep framing tax credits and deductions differently from direct spending (Thuronyi 1988, p. 1172; Zelinsky 2005, p. 826). Perceived reference points are determined by a confluence of psychological, historical, social, and legal factors. It would be a mistake to overstate the role of the last factor.

A more compelling claim rooted in the impact of legal norms on perceived reference points is a *prima facie* argument for legal conservatism. Typically, a change in legal norms yields gains for some and losses to others. Loss aversion entails that the negative impact of a certain decrease of the losers' entitlements is likely to be greater than the positive impact of a comparable increase in the gainers' entitlements. This consideration does not necessarily block legal reforms, yet it calls for caution. *Ceteris paribus*, a legal reform should only be pursued if the gains from changing existing norms outweigh its costs, where gains and losses are weighted differently as the latter loom larger than the former.

Finally, once it is understood that legal norms can sometimes affect people's choices and behavior by setting the reference point, the question arises as to whether and under what circumstances such "manipulation" would be legitimate and desirable. In recent years, this question has been extensively debated under the heading of "soft" or "libertarian" paternalism (Thaler and Sunstein 2009; Camerer et al. 2003; Klick and Mitchell 2006; Glaeser 2010). But the issue of reframing by the law goes far beyond paternalism. Framing of decisions by the law may be used to attain goals other than enhancing

people's own welfare (on legal paternalism, see also Zamir 1998; Zamir and Medina 2010, pp. 313–47). It may be used to maximize overall social utility, as in the case of increasing cadaveric organ donations by presumed-consent legislation. It may also be used to redistribute power or wealth, as in the case of proconsumer or proemployee contractual default rules (Zamir 1997, pp. 1760–62, 1782–84). In general, regulation by reframing decisions has the important advantage of entailing lesser curtailment of people's freedom, compared to most alternatives.

In summary, the potential effect of legal norms on people's perceived reference point is a factor worth considering. Sometimes it militates against changing the legal status quo, because the losses created by such a change loom larger than the gains. However, it can sometimes produce great improvement with very minor interference in people's freedom. In many (and probably most) cases, any attempt to change people's perceptions through the law is likely to be futile. Even when lawmakers do not strive to alter the prevailing framing of decisions, they ought to be aware that legal norms can reinforce this framing—which may or may not be desirable.

## 6.5 Lawmakers' Loss Aversion

Thus far, the discussion has focused on the normative aspects of the loss aversion of people to whom the law applies. Loss aversion presumably also characterizes legal policymakers, such as legislators and judges. Some studies have indeed examined the effects of framing, loss aversion, as well as the related phenomena of status quo bias and the endowment effect on legal decision-making (see chapter 26 by Teichman and Zamir in this volume). No doubt, questionable or objectionable framing of reference points may adversely affect legal policymaking. Cognizant of the powerful effect of reference-dependence, policymakers ought to take heed of manipulative, or even innocuous, framing of decisions. They should be suspicious of strategic framing of (factual, policy, or normative) issues by lobbyists in the legislative process and by advocates in adjudication (as well as by colleagues in collective decision-making bodies). Moreover, conscious framing and reframing of an issue in different ways is often an eye-opening exercise. It may help to challenge old truths and provide an antidote to excessive conservatism. Such reframing has been proposed and debated in various contexts, including affirmative action (Ayres and Vars 1998, pp. 1616–19), and taxation (see section 6.4).

## 7 CONCLUSION

---

Loss aversion is a robust behavioral phenomenon. It sheds light on people's decision-making in various legally pertinent contexts, explains how the law can affect human behavior in subtle ways, and elucidates basic features of the legal system. It carries important normative implications both when the law strives to enhance human

well-being and when it assesses people's behavior in the absence or presence of possible legal norms.

Although the groundbreaking studies of loss aversion are more than thirty years old, the interest in this phenomenon and in related ones, such as the endowment effect, the status quo and omission biases, and the escalation of commitment, is far from decreasing. Dozens of experimental, empirical, and theoretical investigations of these phenomena are constantly carried out by psychologists, economists, and recently jurists as well. While jurists were quick to grasp the importance of these phenomena for legal theory and policymaking (e.g., Kelman 1979), it took some time before they engaged in direct empirical and experimental studies of those phenomena (chapter 5 by Engel in this volume). Along with developments in basic psychological and behavioral-economic studies, it seems that the greatest potential for future advancement in the study of loss aversion and the law lies in such empirical and experimental legal studies. To rephrase one of the analyses described above, conducting such studies seems to be a pure positive gamble, involving potential gains and no losses.

## ACKNOWLEDGMENTS

---

This research was supported by the Israel Science Foundation (grant No. 100/11) and the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation (grant No. 1821/12).

## REFERENCES

---

- Abadie, A., and S. Gay. 2006. The Impact of Presumed Consent Legislation on Cadaveric Organ Donation: A Cross-Country Study. *Journal of Health Economics* 25:599–620.
- Abeler J., A. Falk, L. Goette, and D. Hoffman. 2011. Reference Points and Effort Provision. *American Economic Review* 110:470–92.
- Alchian, A. A. 1953. The Meaning of Utility Measurement. *American Economic Review* 43:26–50.
- Allen, R. J. 1994. How Presumptions Should Be Allocated: Burdens of Proof, Uncertainty, and Ambiguity in Modern Legal Discourse. *Harvard Journal of Law and Public Policy* 17:627–46.
- Anderson, C. J. 2003. The Psychology of Doing Nothing: Forms of Decision Avoidance Result from Reason and Emotion. *Psychological Bulletin* 129:139–67.
- Arkes, H. R., and C. Bloomer. 1985. The Psychology of Sunk Costs. *Organizational Behavior and Human Decision Processes* 35:124–40.
- Arkes, H. R., D. Hirshleifer, D. Jiang, and S. Lim. 2008. Reference Point Adaptation: Tests in the Domain of Security Trading. *Organizational Behavior and Human Decision Processes* 105:67–81.
- Ayres, I., and F. E. Vars. 1998. When Does Private Discrimination Justify Public Affirmative Action? *Columbia Law Review* 98:1577–641.

- Babcock, L., H. S. Farber, C. Fobian, and E. Shafir. 1995. Forming Beliefs about Adjudicated Outcomes: Perceptions of Risk and Reservation Values. *International Review of Law and Economics* 15:289–303.
- Bakos Y., F. Marotta-Wurgler, and D. R. Trossen. 2009. Does Anyone Read the Fine Print? Testing a Law and Economics Approach to Standard Form Contracts. Working paper. Available at <http://ssrn.com/abstract=1443256>.
- Bar-Hillel, M., and E. Neter. 1996. Why Are People Reluctant to Exchange Lottery Tickets? *Journal of Personality and Social Psychology* 70:17–27.
- Baron, J., and I. Ritov. 1994. Reference Points and Omission Bias. *Organizational Behavior and Human Decision Processes* 59:475–98.
- Baumeister, R. F., E. Bratslavsky, C. Finkenauer, and K. D. Vohs. 2001. Bad Is Stronger Than Good. *Review of General Psychology* 5:323–70.
- Becher, S. I. 2007. Behavioral Science and Consumer Standard Form Contracts. *Louisiana Law Review* 68:117–79.
- Becher S. I., and T. Z. Zarsky. 2011. Open Doors, Trap Doors, and the Law. *Law and Contemporary Problems* 74:63–90.
- Bell, D. 1982. Regret in Decision Making under Uncertainty. *Operations Research* 30:961–81.
- Ben-Shahar, O., and J. A. E. Pottow. 2006. On the Stickiness of Default Rules. *Florida State University Law Review* 33:651–82.
- Benartzi, S., and R. H. Thaler. 1995. Myopic Loss Aversion and the Equity Premium Puzzle. *Quarterly Journal of Economics* 110:73–92.
- Birnbaum, M. H. 2008. New Paradoxes of Risky Decision Making. *Psychological Review* 115:463–501.
- Blumenthal, J. A. 2012. Group Deliberation and the Endowment Effect: An Experimental Study. *Houston Law Review* 50:41–70.
- Bokhari, S., and D. Geltner. 2011. Loss Aversion and Anchoring in Commercial Real Estate Pricing: Empirical Evidence and Price Index Implications. *Real Estate Economics* 39:635–70.
- Boonen, L. H. H. M., B. Donkers, and F. T. Schut. 2011. Channeling Consumers to Preferred Providers and the Impact of Status Quo Bias: Does Type of Provider Matter? *Health Services Research* 46:510–30.
- Brennan, S. 1995. Thresholds for Rights. *Southern Journal of Philosophy* 33:143–68.
- Brickman, L. 2003. The Market for Contingent Fee-Financed Tort Litigation: Is It Price Competitive? *Cardozo Law Review* 25:68–128.
- Brockner, J. 1992. The Escalation of Commitment to a Failing Course of Action: Toward Theoretical Progress. *Academy of Management Review* 17:39–61.
- Byun S.-E., and B. Sternquist. 2012. Here Today, Gone Tomorrow: Consumer Reactions to Perceived Limited Availability. *Journal of Marketing Theory and Practice* 20:223–34.
- Camerer, C. F. 2000. Prospect Theory in the Wild: Evidence from the Field. Pp. 288–300 in *Choices, Values, and Frames*, edited by D. Kahneman and A. Tversky. Cambridge: Cambridge University Press.
- Camerer, C. F., L. Babcock, G. Loewenstein, and R. Thaler. 1997. Labor Supply of New York City Cabdrivers: One Day at a Time. *Quarterly Journal of Economics* 112:407–41.
- Camerer, C. F., S. Issacharoff, G. Loewenstein, T. O'Donoghue, and M. Rabin. 2003. Regulation for Conservatives: Behavioral Economics and the Case for “Asymmetric Paternalism.” *University of Pennsylvania Law Review* 151:1211–54.
- Cialdini, R. B. 2009. *Influence: Science and Practice*. 5th ed. Boston: Pearson / Allyn and Bacon.
- Cranston, M. W. 1973. *What Are Human Rights?* New York: Taplinger.

- Cross, F. B. 2001. The Error of Positive Rights. *UCLA Law Review* 48:857–924.
- Currie, D. 1986. Positive and Negative Constitutional Rights. *University of Chicago Law Review* 53:863–90.
- DellaVigna, S. 2009. Psychology and Economics: Evidence from the Field. *Journal of Economic Literature* 47:315–72.
- DeScioli, P., J. Christner, and R. Kurzban. 2011. The Omission Strategy. *Psychological Science* 22:442–46.
- Dittmann, I., E. Maug, and O. Spalt. 2010. Sticks or Carrots? Optimal CEO Compensation When Managers Are Loss Averse. *Journal of Finance* 65:2015–50.
- Donnelly, J. 2003. *Universal Human Rights in Theory and Practice*. 2nd ed. Ithaca, NY: Cornell University Press.
- Eidelman, S., and C. S. Crandall. 2012. Bias in Favor of the Status Quo. *Social and Personality Psychology Compass* 6:270–81.
- Epstein, R. A. 1994. The Ubiquity of the Benefit Principle. *Southern California Law Review* 67:1369–409.
- Ert, E., and I. Erev. 2008. The Rejection of Attractive Gambles, Loss Aversion, and the Lemon Avoidance Heuristic. *Journal of Economic Psychology* 29:715–23.
- Fischer, G. W., M. S. Kamlet, S. E. Fienberg, and D. Schkade. 1986. Risk Preferences for Gains and Losses in Multiple Objective Decision Making. *Management Science* 32:1065–86.
- Frederick, S., and G. Loewenstein. 1999. Hedonic Adaptation. Pp. 302–29 in *Well-Being: The Foundations of Hedonic Psychology*, edited by D. Kahneman, E. Diener, and N. Schwarz. New York: Russell Sage Foundation.
- Friedman, R. D. 1997. Dealing with Evidentiary Deficiency. *Cardozo Law Review* 18:1961–86.
- Fryer, R. G., S. D. Levitt, J. List, and S. Sadoff. 2012. Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment. NBER Working Paper no. 18237. Available at <http://www.nber.org/papers/w18237>.
- Gal, D. 2006. A Psychological Law of Inertia and the Illusion of Loss Aversion. *Judgment and Decision Making* 1:23–32.
- Galini, A. 2013. Endowment Effect in Negotiations: Group versus Individual Decision-Making. *Theory and Decision* 75:389–401.
- Gardbaum, S. 2008. The Myth and the Reality of American Constitutional Exceptionalism. *Michigan Law Review* 107:391–466.
- Gavison, R. 2003. On the Relationships between Civil and Political Rights, and Social and Economic Rights. Pp. 23–55 in *The Globalization of Human Rights*, edited by J.-M. Coicaud et al. Tokyo: United Nations University Press.
- Gilbert, D. T., C. K. Morewedge, J. L. Risen, and T. D. Wilson. 2004. Looking Forward to Looking Backward: The Misprediction of Regret. *Psychological Science* 15:346–50.
- Glaeser, E. L. 2010. Paternalism and Psychology. *University of Chicago Law Review* 73:133–56.
- Goetz, C. J., and R. E. Scott. 1985. The Limits of Expanded Choice: An Analysis of the Interactions between Express and Implied Contract Terms. *California Law Review* 73:261–322.
- Goodman, J. C. 1978. An Economic Theory of the Evolution of the Common Law. *Journal of Legal Studies* 7:393–406.
- Gordon, W. J. 1992. Of Harms and Benefits: Torts, Restitution, and Intellectual Property. *Journal of Legal Studies* 21:449–82.
- Grosskopf, O. 2001. Protection of Competition Rules via the Law of Restitution. *Texas Law Review* 79:1981–2019.



- Guthrie, C. 1999. Better Settle Than Sorry: The Regret Aversion Theory of Litigation Behavior. *University of Illinois Law Review* 1999:43–90.
- Guthrie, C. 2000. Framing Frivolous Litigation: A Psychological Theory. *University of Chicago Law Review* 67:163–216.
- Guthrie, C. 2003. Prospect Theory, Risk Preference, and the Law. *Northwestern University Law Review* 97:1115–63.
- Guthrie, C., and T. E. George. 2005. Judicial Decisionmaking: The Futility of Appeal: Disciplinary Insights into the “Affirmance Effect” on the United States Courts of Appeals. *Florida State University Law Review* 32:357–85.
- Guthrie, C., J. J. Rachlinski, and A. J. Wistrich. 2001. Inside the Judicial Mind. *Cornell Law Review* 86:777–830.
- Haigh, M. S., and J. A. List. 2005. Do Professional Traders Exhibit Myopic Loss aversion? An Experimental Analysis. *Journal of Finance* 60:523–34.
- Hanson, J. D., and D. A. Kysar. 1999. Taking Behavioralism Seriously: The Problem of Market Manipulation. *New York University Law Review* 74:630–749.
- Harinck, F., E. Van-Dijk, I. Van-Beest, and P. Mersmann. 2007. When Gains Loom Larger Than Losses: Reversed Loss Aversion for Small Amounts of Money. *Psychological Science* 18:1099–105.
- Heath, C. 1995. Escalation and De-escalation of Commitment in Response to Sunk Costs: The Role of Budgeting in Mental Accounting. *Organizational Behavior and Human Decision Processes* 62:38–54.
- Hogarth, R. M. 1987. *Judgment and Choice*. Chichester: John Wiley & Sons.
- Holmes, S., and C. R. Sunstein. 1999. *The Cost of Rights: Why Liberty Depends on Taxes*. New York: W. W. Norton.
- Horowitz, J. K., and K. E. McConnell. 2002. A Review of WTA/WTP Studies. *Journal of Environmental Economics and Management* 44:426–47.
- Inman, J. J., and L. McAlister. 1994. Do Coupon Expiration Dates Effect Human Behavior? *Journal of Marketing Research* 31:423–28.
- Johnson, E. J., and D. Goldstein. 2003. Do Defaults Save Lives? *Science* 302:1338–39.
- Johnson, S. R. 1999. The Dangers of Symbolic Legislation: Perceptions and Realities of the New Burden-of-Proof Rules. *Iowa Law Review* 84:413–97.
- Kagan, S. 1989. *The Limits of Morality*. Oxford: Oxford University Press.
- Kagan, S. 1998. *Normative Ethics*. Boulder, CO: Westview Press.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1990. Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy* 98:1325–48.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1991. The Endowment Effect, Loss Aversion, and the Status Quo Bias. *Journal of Economic Perspectives* 5:193–206.
- Kahneman, D., and D. T. Miller. 1986. Norm Theory: Comparing Reality to Its Alternatives. *Psychological Review* 93:136–53.
- Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47:263–91.
- Kahneman, D., and A. Tversky. 1984. Choices, Values, and Frames. *American Psychologist* 39:341–50.
- Kamm, F. M. 1992. Non-consequentialism, the Person as an End-in-Itself, and the Significance of Status. *Philosophy and Public Affairs* 21:354–89.
- Kelman, M. 1979. Consumption Theory, Production Theory, and Ideology in the Coase Theorem. *Southern California Law Review* 52:669–98.



- Kermer, D. A., E. Driver-Linn, T. D. Wilson, and D. T. Gilbert. 2006. Loss Aversion Is an Affective Forecasting Error. *Psychological Science* 17:649–53.
- Kleinbard, E. D. 2010. Tax Expenditure Framework Legislation. *National Tax Journal* 63:353–81.
- Klick J., and G. Mitchell. 2006. Government Regulation of Irrationality: Moral and Cognitive Hazards. *Minnesota Law Review* 90:1620–63.
- Knetsch, J. L. 1989. The Endowment Effect and Evidence of Nonreversible Indifference Curves. *American Economic Review* 79:1277–84.
- Knetsch, J. L. 1992. Preferences and Nonreversibility of Indifference Curves. *Journal of Economic Behavior and Organization* 17:131–39.
- Knetsch, K. L., and J. A. Sinden. 1984. Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value. *Quarterly Journal of Economics* 99:507–21.
- Korobkin, R. 1998. The Status Quo Bias and Contract Default Rules. *Cornell Law Review* 83:608–87.
- Korobkin, R. B., and C. Guthrie. 1994. Psychological Barriers to Litigation Settlement: An Experimental Approach. *Michigan Law Review* 93:107–92.
- Korobkin, R. B., and C. Guthrie. 1997. Psychology, Economics, and Settlement: A New Look on the Role of the Lawyer. *Texas Law Review* 76:77–141.
- Korobkin, R. B., and T. S. Ulen. 2000. Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics. *California Law Review* 88:1051–144.
- Kőszegi, B., and M. Rabin. 2006. A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics* 121:1133–65.
- Kőszegi, B., and M. Rabin. 2007. Reference-Dependent Risk Attitude. *American Economic Review* 97:1047–73.
- Kritzer, H. M. 1990. *The Justice Broker: Lawyers and Ordinary Litigation*. New York: Oxford University Press.
- Kritzer, H. M. 2002. Seven Dogged Myths Concerning Contingency Fees. *Washington University Law Quarterly* 80:739–94.
- Ku, G. 2008. Learning to De-escalate: The Effects of Regret in Escalation of Commitment. *Organizational Behavior and Human Decision Processes* 105:221–32.
- Langevoort, D. C. 1996. Selling Hope, Selling Risk: Some Lessons for Law from Behavioral Economics about Stockbrokers and Sophisticated Customers. *California Law Review* 84:627–701.
- Larrick, R. P., and T. L. Boles. 1995. Avoiding Regret in Decisions with Feedback: A Negotiation Example. *Organizational Behavior and Human Decision Processes* 63:87–97.
- Levmore, S. 1985. Explaining Restitution. *Virginia Law Review* 71:65–124.
- Lewinsohn-Zamir, D. 2003. The Objectivity of Well-Being and the Objectives of Property Law. *New York University Law Review* 78:1669–754.
- Llewellyn, K. N. 1937. On Warranty of Quality, and Society: II. *Columbia Law Review* 37:341–409.
- Loewenstein, G., and D. Adler. 1995. A Bias in the Prediction of Tastes. *Economic Journal* 105:929–37.
- Loewenstein, G. F., L. Thompson, and M. H. Bazerman. 1989. Social Utility and Decision Making in Interpersonal Contexts. *Journal of Personality and Social Psychology* 57:426–41.
- Loomes, G., and R. Sugden. 1982. Regret Theory: An Alternative Theory of Rational Choice under Uncertainty. *Economic Journal* 92:805–24.

- Madrian, B. C., and D. F. Shea. 2001. The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior. *Quarterly Journal of Economics* 116:1149–87.
- Markowitz, H. 1952. The Utility of Wealth. *Journal of Political Economy* 60:151–58.
- McCarthy, A. M., F. D. Schoorman, and A. C. Cooper. 1993. Reinvestment Decisions by Entrepreneurs: Rational Decision-Making or Escalation of Commitment? *Journal of Business Venturing* 8:9–24.
- McCormick, C. T. 2006. *McCormick on Evidence*. Vol. 2. 6th ed., edited by K. S. Broun et al. St. Paul, MN.: Thompson/West.
- McGraw, A. P., J. T. Larsen, D. Kahneman, and D. Schkade. 2010. Comparing Gains and Losses. *Psychological Science* 21:1438–45.
- McKenzie, C. R. M., M. J. Liersch, and S. R. Finkelstein. 2006. Recommendations Implicit in Policy Defaults. *Psychological Science* 17:414–20.
- Mehra, R., and E. C. Prescott. 1985. The Equity Premium: A Puzzle. *Journal of Monetary Economics* 15:145–61.
- Mellers, B. A., and I. Ritov. 2010. How Beliefs Influence the Relative Magnitude of Pleasure and Pain. *Journal of Behavioral Decision Making* 23:369–82.
- Mellers, B. A., A. Schwartz, K. Ho, and I. Ritov. 1997. Decision Affect Theory: Emotional Reactions to the Outcomes of Risky Options. *Psychological Science* 8:423–29.
- Mellers, B. A., A. Schwartz, and I. Ritov. 1999. Emotion-Based Choice. *Journal of Experimental Psychology: General* 128:332–45.
- Moshinsky, A., and M. Bar-Hillel. 2010. Loss Aversion and the Status Quo Label Bias. *Social Cognition* 28:191–204.
- Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Parisi, F. 2004. The Efficiency of the Common Law Hypothesis. Pp. 195–98 in *The Encyclopedia of Public Choice*, vol. 2, edited by C. K. Rowley and F. Schneider. Dordrecht: Kluwer Academic Publishers.
- Plott, C. R., and K. Zeiler. 2005. The Willingness to Pay–Willingness to Accept Gap, the “Endowment Effect,” Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. *American Economic Review* 95:530–45.
- Plott, C. R., and K. Zeiler. 2007. Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory? *American Economic Review* 97:1449–66.
- Porat, A. 2009. Private Production of Public Goods: Liability for Unrequested Benefits. *Michigan Law Review* 108:189–227.
- Prentice, R. A., and J. J. Koehler. 2003. A Normality Bias in Legal Decision Making. *Cornell Law Review* 88:583–650.
- Priest, G. L. 1977. The Common Law Process and the Selection of Efficient Rules. *Journal of Legal Studies* 6:65–82.
- Rachlinski, J. J. 1996. Gains, Losses, and the Psychology of Litigation. *Southern California Law Review* 70:113–85.
- Rachlinski, J. J., and F. Jourden. 1998. Remedies and the Psychology of Ownership. *Vanderbilt Law Review* 51:1541–82.
- Ritov, I., and J. Baron. 1990. Reluctance to Vaccinate: Omission Bias and Ambiguity. *Journal of Behavioral Decision Making* 3:263–77.
- Ritov, I., and J. Baron. 1992. Status-Quo and Omission Biases. *Journal of Risk and Uncertainty* 5:49–61.
- Ritov, I., and J. Baron. 1995. Outcome Knowledge, Regret, and Omission Bias. *Organizational Behavior and Human Decision Processes* 64:119–27.

- Rozin, P., and E. B. Royzman. 2001. Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review* 5:296–320.
- Rubin, P. H. 1977. Why Is the Common Law Efficient? *Journal of Legal Studies* 6:51–63.
- Rubin, P. H. 2005. Micro and Macro Legal Efficiency: Supply and Demand. *Supreme Court Economic Review* 13:19–34.
- Rubinstein, A. 2006. *Lecture Notes in Microeconomic Theory*. Princeton, NJ: Princeton University Press.
- Samuelson, P. A., and W. D. Nordhaus. 2010. *Economics*. 19th ed. Boston: McGraw-Hill Irwin.
- Samuelson, W., and R. Zeckhauser. 1988. Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty* 1:7–59.
- Santore, R., and A. D. Viard. 2001. Legal Fee Restrictions, Moral Hazard, and Attorney Rents. *Journal of Law and Economics* 44:549–72.
- Savage, L. J. 1972. *The Foundations of Statistics*. 2nd ed. New York: Dover.
- Sayman, S., and A. Öncüler. 2005. Effects of Study Design Characteristics on the WTA-WTP Disparity: A Meta Analytical Framework. *Journal of Economic Psychology* 26:289–312.
- Scheffler, S. 1988. Introduction. Pp. 1–13 in *Consequentialism and Its Critics*, edited by S. Scheffler. Oxford: Oxford University Press.
- Schweitzer, M. 1994. Disentangling Status Quo and Omission Effects: An Experimental Analysis. *Organizational Behavior and Human Decision Processes* 58:457–76.
- Smith, G. E., and T. T. Nagle. 1995. Frames of Reference and Buyers' Perceptions of Price and Value. *California Management Review* 38:98–116.
- Spier, K. E. 2007. Litigation. Pp. 259–342 in *Handbook of Law and Economics*, vol. 1, edited by A. M. Polinsky, and S. Shavell. Amsterdam: Elsevier.
- Stake, J. E. 1991. Status and Incentive Aspects of Judicial Decisions. *Georgetown Law Journal* 79:1447–97.
- Staw, B. M. 1976. Knee-Deep in the Big Muddy: A Study of Escalating Commitment to a Chosen Course of Action. *Organizational Behavior and Human Decision Processes* 16:27–44.
- Staw, B. M., and J. Ross. 1989. Understanding Behavior in Escalation Situations. *Science* 246:216–20.
- Stein, A. 1998. An Essay on Uncertainty and Fact-Finding in Civil Litigation, with Special Reference to Contract Cases. *University of Toronto Law Journal* 48:299–351.
- Sugarman, D. B. 1986. Active versus Passive Euthanasia: An Attributional Analysis. *Journal of Applied Social Psychology* 16:60–76.
- Sunstein, C. R. 2005. Why Does the American Constitution Lack Social and Economic Guarantees? Pp. 90–110 in *American Exceptionalism and Human Rights*, edited by M. Ignatieff. Princeton, NJ: Princeton University Press.
- Surrey S. S., and P. R. McDaniel. 1985. *Tax Expenditures*. Cambridge, MA: Harvard University Press.
- Thaler, R. H. 1980. Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior and Organization* 1:39–60.
- Thaler, R. H., and E. J. Johnson. 1990. Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. *Management Science* 36:643–60.
- Thaler, R. H., and C. R. Sunstein. 2009. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Rev. ed. London: Penguin.
- Thuronyi, V. 1988. Tax Expenditures: A Reassessment. *Duke Law Journal* 1988:1155–206.
- Tversky, A., and D. Kahneman. 1991. Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics* 106:1039–61.

- Tversky, A., and D. Kahneman. 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* 5:297–323.
- Vars, F. E. 1996. Attitudes toward Affirmative Action: Paradox or Paradigm? Pp. 73–100 in *Race versus Class: The New Affirmative Action Debate*, edited by C. M. Swain. Lanham, MD: University Press of America.
- Zamir, E. 1997. The Inverted Hierarchy of Contract Interpretation and Supplementation. *Columbia Law Review* 97:1710–803.
- Zamir, E. 1998. The Efficiency of Paternalism. *Virginia Law Review* 84:229–86.
- Zamir, E. 2012. Loss Aversion and the Law. *Vanderbilt Law Review* 65:829–94.
- Zamir, E., and B. Medina. 2010. *Law, Economics, and Morality*. New York: Oxford University Press.
- Zamir, E., and I. Ritov. 2010. Revisiting the Debate over Attorneys' Contingent Fees: A Behavioral Analysis. *Journal of Legal Studies* 39:245–88.
- Zamir, E., and I. Ritov. 2012. Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation. *Journal of Legal Studies* 41:165–207.
- Zelinsky, E. A. 2005. Do Tax Expenditures Create Framing Effects? Volunteer Firefighters, Property Tax Exemptions, and the Paradox of Tax Expenditures Analysis. *Virginia Tax Review* 24:797–830.

## CHAPTER 12

---

# WRESTLING WITH THE ENDOWMENT EFFECT, OR HOW TO DO LAW AND ECONOMICS WITHOUT THE COASE THEOREM

---

RUSSELL KOROBKIN

STANDARD economic analysis of law, dating to the publication of Ronald Coase's *The Problem of Social Cost* (1960), assumes that the initial allocation of legal rights will not affect the final allocation of those rights in the absence of transaction costs. To borrow one of Coase's own examples, whether the law protects Farmer's land from trespass or recognizes neighboring Rancher's right to freely graze his cattle will not affect how the land is put to use. If the law protects Farmer but the land is more valuable as pasture than it is for growing crops, Rancher will pay Farmer for the right to graze his cattle. If the law protects Rancher but the land is more valuable for crops than pasture, Farmer will pay Rancher to pen his cattle. The assignment of the legal entitlement will affect the distribution of wealth (Farmer will be richer if he has the right to exclude), but not whether the land is put to its most efficient use.

A crucial premise of this assumption is that the value actors place on legal entitlements is independent of legal title or physical possession. The value to Farmer of the use of his fields for crops might be high or low, but it is the same whether he has the right or ability to exclude Rancher's herd (and, thus, he might consider selling grazing rights) or Rancher has the right or ability to graze (and, thus, Farmer might consider paying Rancher to pen his cattle). Whether or not this premise is correct turns out to have significant implication for legal policy across substantive fields, from how the state should initially allocate entitlements, whether it should reallocate entitlements, how it should regulate the consensual exchange of entitlements, to how it should protect and vindicate entitlement ownership.

The empirical evidence demonstrates, however, that the context surrounding entitlements can affect their valuation. Research on the "status quo bias" (Samuelson and Zeckhauser 1988) indicates that individuals tend to prefer the present state of the world to alternative states, all other things being equal. The term "endowment effect" (Thaler

1980) stands for the slightly narrower principle that people tend to value entitlements more when they are endowed with them than when they are not (but is often used interchangeably with “status quo bias”). A consequence of the endowment effect is an “offer-asking” gap (Kennedy 1981), meaning that the same person will often demand a higher price to sell an entitlement that is part of his endowment than he will pay for the same entitlement if it is not part of his endowment.

This body of research has spurred a great deal of analysis and debate in the legal academic literature. Prior to 1990, only seventeen law journal articles available in the Westlaw database contained either of the terms “endowment effect” or “status quo bias.” By the end of 2000, 358 total articles met this requirement. As of the end of 2012, the number had ballooned to 1,619.<sup>1</sup> Legal scholars have universally grasped the most important positive implication of the endowment effect—that legal entitlements will not change hands as often in the free market as the Coase Theorem—and thus traditional law and economics—assumes. Scholars have also attempted to revisit arguments about normatively appropriate legal policy in light of the endowment effect. This effort has been made with varying levels of success and sophistication, and continued progress on this score is critical for the continued development of the field of behavioral law and economics.

This chapter has two purposes. First, it describes the endowment effect, with attention not only to what we know about it, but also what remains unclear about both its scope and its underlying causal mechanism. Second, it demonstrates how the endowment effect is important to a wide range of substantive legal topics, but that, while its positive implications for law are often straightforward, its normative implications have proven to be more elusive.

Section 1 presents the core empirical findings concerning the endowment effect, and Section 2 evaluates the evidence for several highly contested interpretations of what psychological process or processes cause the endowment effect. Section 3 considers examining how the endowment effect might bear on positive and normative issues in four broad categories of law: the initial assignment of entitlements, the potential reassignment of entitlements, the facilitation of private transfers of entitlements, and the protection of entitlements through the judicial system. This part demonstrates that, used cautiously and judiciously, evidence of the endowment effect has the potential to sharpen normative legal policy analysis, but also that this enterprise is complicated and fraught with peril.

## 1 EVIDENCE OF THE ENDOWMENT EFFECT

---

### 1.1 From Contingent Valuation to Controlled Experiments

Studies of the value individuals place on public goods, often environmental protection, conducted in the late 1960s and 1970s suggested that possession of legal rights affects

<sup>1</sup> All of these counts are based on computerized searches of the Westlaw “Journals and Law Reviews” (JLR) database.

valuations. Most of these results came from surveys in which researchers asked some respondents how much they would be willing to pay for a public good and ask others how much they would demand to give up a public good, assuming it were provided. In one example, 2,000 duck hunters were surveyed about the value they would place on protecting a wetland from development. Hunters reported that they would pay, on average, \$247 per person per season for the right to prevent development in order to make hunting viable, while reporting an average demand of \$1,044 dollars each to give up an entitlement to protect the area for hunting (Hammack and Mallard Brown 1974).

This type of “contingent valuation” survey is subject to the objection that subjects lacked any real incentive to provide truthful or well-considered answers, so researchers began to employ experimental tests that provide subjects with clear personal incentives to both think carefully about and reveal their true valuations. The simpler of these present subjects with the option of making trades, whereas the more complex require subjects to price entitlements.

In trading experiments, experimenters divide subjects into two groups and either give members of one group a good or an approximately equivalent amount of cash or give members of the two groups different goods that are approximately equivalent in value. Subjects are then given the opportunity to trade their endowment to the experimenter for the alternative item. The experiments are designed so that the transaction costs of trading approach zero, so the Coase Theorem implies that the distribution of goods among subjects at the end of the experiment should be the same for each group. If final endowments are biased by initial endowments, in contrast, this result suggests an endowment effect. Jack Knetsch produced the most famous of the trading experiments. Knetsch and Sinden (1984) endowed some experimental subjects with lottery tickets for a \$50 cash prize and gave others \$3 in cash. They then offered to buy or sell tickets for \$3. Only 38% of the non-ticket holders were willing to trade their \$3 for a ticket, but 82% of the subjects originally endowed with tickets chose to hold on to them. Knetsch (1989) then endowed some with coffee mugs and others with large chocolate bars and offered to swap objects with all subjects. Only 11% of the mug owners traded their mugs for a chocolate bar, and only 10% of the chocolate bar owners swapped for a mug.

In pricing experiments, experimenters elicit the precise willingness to pay (“WTP”) and willingness to accept (“WTA”) values of subjects as in hypothetical surveys, but they create miniature markets in which the subjects maximize their personal utility by truthfully revealing their valuations. In the most famous pricing experiment, Kahneman, Knetsch, and Thaler (1990) (KKT) provided one-half of their subjects with a coffee mug bearing the Cornell University logo. The experimenters told the subjects who received the mug that they would have an opportunity to sell it, and they told the remainder of subjects that they would have an opportunity to purchase one of the mugs. The experimenters then (1) asked mug holders (“sellers”) to fill out a form indicating, for each 50-cent interval between 25 cents and \$8.75, whether they would prefer to keep the mug or to sell it at that price and (2) asked non-mug holders (“buyers”) to indicate whether they would be willing to buy a mug at each of the same price intervals. The experimenters explained in advance that, based on the responses, they would determine the market price of the mugs. All sellers who indicated on their form



a willingness to sell at that price or lower would sell for the market price and buyers who indicated a willingness to buy at that price or higher would buy a mug at the market price. Because exchanges would take place at the market-determined price, which would be set by all of the subject responses, buyers lacked a strategic incentive to understate their true valuations and sellers lacked an incentive to exaggerate theirs.

In four iterations of the experiment with the same subjects, buyers provided a median WTP of \$2.25 to \$2.75, and sellers provided a median WTA of \$5.25 each time. The experimenters achieved nearly identical results using pens as a commodity rather than mugs, finding that sellers' WTA values were approximately twice buyers' WTP values, and that only a fraction of the predicted number of trades between buyers and sellers actually took place. The authors found similar results in a follow-up experiment in which trading prices were selected at random rather than based on subject bids (which, in theory, could have caused some subjects to believe they were best off trying to "buy low" and "sell high" rather than reveal their true valuations).

Hundreds of experiments demonstrating the endowment effect (Georgantzis and Navarro-Martinez 2010) show that early findings were not an artifact of experimenters' use of hypothetical surveys, in which respondents have no personal stake in truthfully revealing their valuations. A study of forty-five early endowment effect studies found, in fact, that there is no significant difference in the size of the effect (i.e., the ratio between WTA and WTP valuations) between real experiments and hypothetical questionnaires (Horowitz and McConnell 2002).

## 1.2 What Is the Reference Point?

Evidence of an endowment effect raises the obvious question of what conditions will cause individuals to perceive their status as "endowed" or "unendowed," and act accordingly. That is, what is the reference point that determines whether individuals view a potential transaction as one that, if consummated, would result in them gaining an entitlement rather than as one that would involve losing an entitlement?

Although the foundational endowment effect experiments concern tangible consumer items, such as mugs, pens, chocolate bars, or lottery tickets, other experiments have shown that the endowment effect exists for less tangible legal entitlements that ordinary people might not conceptualize as property rights, at least in some circumstances. One study, for example, found a large differential between subjects' WTP to avoid having to taste a bad-tasting, but harmless, liquid and their WTA to taste the liquid when they had the right not to taste it (Coursey et al. 1987). Another demonstrated that the endowment effect exists for the right to enjoy clean air, with subjects reporting a willingness to pay less than \$5 per month, on average, to maintain 75 miles of air visibility in the face of potential pollution, but would demand more than \$24 per month to agree to give up a right to avoid increased pollution (Rowe et al. 1980).

Research also suggests that the endowment effect can exist when no actual legal entitlement is at issue at all but subjects expect a certain state of the world. Hartman, Doane, and Woo (1991) found that power company customers who had experienced excellent

service would demand payments to accept a reduction in service quality that were 400% of the amount customers who were accustomed to poorer service would pay for better service. Dubourg, Jones-Lee, and Loomes (1994) found that subjects would demand far more money to agree to purchase a car with one fewer safety feature than the standard model than they would be willing to pay for a car with one more safety feature than the standard model. The results of a famous natural experiment also shows that default expectations can affect revealed valuations, at least in some circumstances. New Jersey and Pennsylvania both enacted laws in the 1980s giving drivers the option of purchasing low-cost insurance with limited rights to sue other drivers in addition or more expensive insurance with greater legal rights. In New Jersey the default option was the limited rights policy, and the broader policy required opting in. In Pennsylvania, the default option was the expansive policy. Large majorities of drivers in both states accepted the default choices rather than opting for the alternative (Cohen and Knetsch 1992).

In a slight variation, Koszegi and Rabin (2006) theorize that the typical reference point is neither legal entitlement nor status quo state of the world, but rather expectations about the future state of the world. Tests of this hypothesis have produced equivocal results. Ericson and Fuster (2010) gave subjects a mug and told half that they would have a 90% chance of having an option later to trade the mug for a pen and the other half that they would have a 10% chance of having that option. When subjects were later given the opportunity to trade goods, those in the first group were significantly more likely to do so than those in the second, suggesting that the expectation of having to keep the mug produced a higher mug valuation than the expectation that the mug could be easily traded. Heffetz and List (2011), in contrast, found that subjects who were told their gift for participating in an experiment would be a mug (or pen) with 99% probability and there was a 1% probability that they could choose between mug and pen were no more likely to prefer the mug (or pen) than subjects who were told that they had a 99% probability of being allowed to choose and a 1% probability of being assigned the mug (or pen).

Uncertainty about the impact of the endowment effect abounds when multiple indicators of entitlement conflict with one another. If your neighbor allows his tree to block your view, in violation of the local zoning ordinance, and you would like him to cut it down, it is unclear who is likely to feel endowed: you possess a legal right, but he possess the physical entitlement (i.e., the tree exists). If your business associate breaches a contract to deliver goods to you, she possess the goods but you have a legal right to performance. It seems likely, in such situations, that the perception of endowment will depend on the salience of other contextual features, and that it is possible that both might feel endowed for purposes of valuation.

### 1.3 Comparability and Certainty of Alternative Endowments

The more difficult it is for individuals to compare two items in a proposed trade, the larger the endowment effect tends to be. Van Dijk and van Knippenberg (1998) gave subjects one of two different bottles of wine as compensation for participating in a study and were given the opportunity to trade. When both bottles of wine were from

the same country, approximately one-half of the subjects made trades, but fewer than one-third made trades when the bottles were from different countries. Bar-Hillel and Neter (1996) found that nearly all subjects endowed with a pen were willing to trade it for an identical pen plus a small cash payment, but most subjects endowed with a lottery ticket refused to trade it for a different ticket (with a different number) for the same lottery plus the small cash payment, presumably because the difference in value between the tickets was potentially substantial *ex post*, whereas there was no possibility that the pens would differ in utility.

Experiments have also found that the endowment effect is more robust for entitlements with no close market substitutes than for goods that have close substitutes or are themselves readily purchasable. In one study, Shogren and coauthors (1994) found no differential between subjects' WTP and WTA for an ordinary candy bar. In contrast, they found an extremely large difference between subjects' WTP to trade a lunch from a local source for a lunch that had been screened for food-borne pathogens and had a lower than average probability of causing food-borne illness and their WTA to trade the screened lunch for the ordinary one. When an entitlement has a close or identical market substitute, such as a candy bar, it is presumably easier for individuals to compare its value to the value of money and determine with a high degree of confidence whether they would be better off with money or the entitlement. At the extreme end of the spectrum are tokens with a fixed redemption value. In their original pricing experiments, KKT found no endowment effect for such tokens—subjects were willing to buy or sell them for almost exactly their exchange value.

There is also some evidence for the subtler hypothesis that the endowment effect will be smaller when a good has a somewhat close substitute as compared to no substitute at all. Adamowicz, Bhardwaj, and Macnab (1993) (ABM) tested subjects' WTP and WTA values for tickets to a National Hockey League playoff game. In one condition, subjects were told that the game would be broadcast live on TV and radio, whereas subjects in another condition were told that the game could be viewed only in person. Although an endowment effect was apparent in both conditions, the effect was substantially larger for subjects who were told that the game could be seen only in person—that is, those for which there was no substitute at all for the tickets.

A meta-analysis has validated the hypothesis that the strength of the endowment effect depends on how close a substitute there is for the entitlement at issue (Horowitz and McConnell 2002). The researchers conclude that, even accounting for differences in study design, the ratio of WTA to WTP “is highest for public and non-market goods, next highest for ordinary private goods, and lowest for experiments involving forms of money” (p. 426).

## 1.4 Market Participation and Orientation to Transactions

If the extent of the endowment effect depends on the existence of close substitutes, it follows that the effect will be larger when individuals evaluate an entitlement for its use value rather than its exchange value, all other things being equal, because an item held

only for its exchange value has ready substitutes. For example, if ABM's hockey ticket subjects assumed that they would sell the tickets to a broker, their WTA and WTP prices in the experiment would likely converge, especially if they knew with certainty what price brokers would pay for such tickets.

This hypothesis finds support, at least inferentially, in two sets of experiments. Arlen, Spitzer, and Talley (2002) found a significant endowment effect for mugs in a pilot experiment designed to mimic the studies of KKT, but not when the subjects were told that the mugs were a factor of production that would affect the profits of their firm and that their salary would be based on the firm's profits. The authors theorize that the context of the experiment likely made salient the value of the mug as a generator of company profits rather than as a unique consumption good, and thus caused subjects to treat the mug as more fungible. Harless (1989) had a group of subjects participate as both buyers and sellers in a series of transactions involving lottery tickets and found a small but non-statistically-significant gap between the median values provided for the two measures. A likely explanation for the lack of significant results is that the opportunity to participate in identical transactions as both buyer and seller focused subjects' attention on the exchange value of the lotteries.

Some studies have found that people who regularly participate in particular types of transactions are less likely to exhibit an endowment effect than infrequent market participants. List (2003, 2011) found that baseball memorabilia collectors who engage in frequent transactions traded one memorabilia item for another of similar desirability frequently enough to suggest they exhibited no endowment effect, but the much more reticent behavior of infrequent traders did demonstrate an endowment effect. The frequent traders who were memorabilia dealers were no more likely to trade than active hobbyists. General experience with market transactions does not appear to reduce the endowment effect, however. Harbaugh, Krause, and Vesterlund (2001) found no significant differences between kindergarten, third-grade, fifth-grade, and college student participants in a simple exchange experiment.

Evidence is mixed on a subject of particular importance to lawyers: Is the endowment effect as strong for agents as for principals involved in a transaction? Marshall, Knetsch, and Sinden (1986) (MKS) answer in the negative. In a series of hypothetical scenarios and one actual choice experiment, they asked one group of subjects whether or not they would pay a fixed amount of money (\$1 or \$2) for a ticket in a lottery with a specified prize, or whether or not they would give up such a ticket for the cash payment. They asked members of a second group whether or not they would advise a friend to buy the ticket or sell the ticket. Subjects acting as principals were much more likely to prefer the tickets to the cash if they were sellers than if they were buyers, whereas there was no significant difference based on role for "advisor" subjects. In contrast, I found a pronounced endowment effect in hypothetical scenarios that asked law students to assume the role of a lawyer and make transactional decisions on behalf of a corporate client: subjects placed a higher value on a contract term that would benefit the client if the term was associated with the status quo contract language than if not (Korobkin 1998a, 1998b).

The difference in results perhaps is explainable by the difference in instructions provided to the experimental subjects. MKS asked their agent subjects to provide advice to a third party, with no apparent consequences for the content of the advice given. My subjects were told that their future opportunities with their client depended on the results of the contracts they negotiated, and the subjects were reminded to make decisions that are in the best interest of the client and will “reflect[] well on [the subject’s] judgment and ability” (Korobkin 1998a, appendix A2, p. 678). This might have caused my subjects to behave more like principals with a personal interest in the transaction and less like disinterested third parties, at least relative to the MKS subjects. A plausible hypothesis is that whether agents exhibit the endowment effect expected of principals depends on the degree to which the incentives of principal and agent are aligned.

## 1.5 Could the Endowment Effect Be an Artifact of Poor Laboratory Practices?

Is it possible that the wealth of experimental evidence that endowment status affects valuations reflects a lack of careful controls in laboratory studies, rather than real phenomena concerning preferences? Plott and Zeiler (2005; 2007) (PZ) contend that the endowment effect, as apparently established with pricing and trading experiments, can be eliminated by employing procedures that protect against subjects’ misunderstanding the nature of the experiments. The implication of their claim is that the endowment effect is unlikely to exist outside the confines of the laboratory.

### 1.5.1 *Pricing Experiments*

In pricing experiments, there is a fear that experimental subjects (or at least some of them) might not understand that it is in their interest to reveal their true valuations of the entitlements being bought and sold, rather than attempting to implement the common bargaining strategy of buying low and selling high. Consequently, reports of high WTA or low WTP values in experiments might reflect (poor) attempts at strategic behavior rather than differences in valuations based on endowment status.

PZ (2005) replicated the KKT mug experiment, but they provided subjects with more thorough training on how to optimally bid, including several specific examples, and an opportunity for subjects to ask questions about the experimental procedures. In addition, the experimenters were blind to WTA and WTP price provided by subjects. They also had two-thirds of their subjects participate in 14 “paid practice rounds,” in which subjects both bought and sold rights to participate in small cash lotteries, prior to participating in the mug experiment, in an effort providing subjects with a substantial opportunity to learn firsthand the relationship between the valuations they provided and the subsequent payouts. In their ultimate mug experiment, PZ found no significant differences between average WTA and WTP prices (in fact, the average WTP was somewhat higher, but not significantly so, than the average WTA). From this finding,

PZ concluded that observed gaps between WTA and WTP prices “are symptomatic of subjects’ misconceptions about the nature of the experimental task,” rather than true differences in valuations.

Although it is possible that PZ’s results show that KKT’s subjects (and those in many other similar experiments) failed to reveal their true valuations of mugs because they did not understand the preference elicitation mechanism while PZ’s subjects uniquely did understand the design mechanism, two features of the original mug experiment make this seem unlikely. First, if subjects incorrectly believed they could benefit by attempting to buy low and sell high, it should not have mattered whether the entitlement at issue had a substitute with a clear value. But unlike in the case of mugs, KKT’s subjects exhibited no endowment effect when they bought and sold tokens exchangeable for cash.

Second, if subjects misunderstood the experimental incentives, experience with the experimental design and outcomes should have reduced or eliminated the endowment effect. But KKT elicited WTP and WTA prices from subjects and constructed markets four consecutive times for both mugs and pens, using the same subjects in the same roles, specifically to test for whether WTP and WTA would converge as subjects became familiar with the experimental process. They found that reported WTP and WTA values and the number of trades that took place between those endowed with the goods and those not endowed were constant across all iterations. KKT concluded that “[o]bserved volume did not increase over successive periods in either the mug or the pen markets, providing no indications that subjects learned to adopt equal buying and selling prices” (KKT 1990, p. 1332). The results were virtually identical whether subjects were told that the market price would be established by constructing supply and demand curves from the subjects’ WTP and WTA prices or that the market price would be determined randomly.

Although PZ report no endowment effect for mugs, they did find an endowment effect across the 14 “practice rounds,” in which they paid subjects real money depending on their WTA and WTP values for participating in lotteries. PZ did not report this fact,<sup>2</sup> but Isoni, Loomes, and Sugden (2011) (ILS) did after obtaining PZ’s original data. ILS replicated this result, and rather pointedly concluded that, in light of a persistent and nondecaying endowment effect for lotteries in both their and PZ’s experiments, it is “not credible to propose that misconceptions about a common set of elicitation procedures” could have persisted through 14 rounds of paid lottery tasks in the PZ experiment and then “suddenly disappear[ed] when the mug task is faced” (ILS 2011, p. 1005). These lottery results do seem to severely undermine PZ’s claim that the endowment effect is really nothing more than the consequence of experimental designs that befuddle subjects.

<sup>2</sup> PZ stated in a footnote that their lottery rounds were “contaminated” because WTA values were elicited in early lottery rounds and WTP in later rounds, and because experimenters publicly answered subjects’ question during these rounds (PZ 2005, 539 n. 15).



What then could explain why PZ (and ILS) found a persistent endowment effect for lotteries but not for mugs? ILS point out two differences between the experimental designs employed in PZ's and KKT's mug experiments. First, PZ placed mugs directly in front of both buyers and sellers, whereas KKT placed mugs only in front of buyers. Second, PZ paid nonowners a "show up fee," which they could of course use to buy a mug, whereas KKT did not. Either or both of these features could have weakened the endowment effect in a way not typical of "real world" situations, in which the distinction between endowed and unendowed status is generally clear and there is no "house money" to back up WTP prices. Another subtle difference is that PZ "trained" subjects to determine WTP values by starting with a low number and then moving up until they reached their valuation and to determine WTA values by starting with a high number and then moving down. Then, after subjects recorded their valuation, PZ prompted them to reconsider whether they had recorded their actual nonstrategic value and change their valuation if necessary. This could have created a demand effect, causing buyers to believe they should move higher and sellers to believe they should move lower. Kingsley and Brown (2012) replicated PZ's training method and PZ's prompt for subjects "to reconsider" and found that preprompt responses produced an endowment effect that then disappeared after postprompt reconsideration.

### 1.5.2 *Trading Experiments*

In any event, if the endowment effect actually results purely from subject confusion concerning valuation elicitation procedures, how can the results of simpler trading experiments—such as the early experiments by Knetsch in which subjects given one item may trade it for another—be explained? PZ (2007) contend that those experiments are not clean tests of the endowment effect because when experimenters give subjects a particular item, the subjects might either infer a signal of high quality or perceive the given item as a "gift" that carries with it sentimental value. In either case, any observed resistance to trading could reflect differences other than endowment status.

To test their conjecture, PZ conducted a trading experiment using mugs and pens. They found that they could eliminate the endowment effect by (a) telling subjects that whether they received a mug or a pen was determined by a coin flip, (b) taking the endowed items away from subjects after providing an opportunity for examination, (c) giving subjects the alternative item for inspection as well, and, finally, (d) asking subjects to indicate which item they wished to "keep" on a decision form.

Critics of this design might fear that it not only removes any hint of signal or gift, it also lessens the psychological indicia of endowment status likely to operate in real-world situations. In response to this potential objection, PZ conducted a follow-up experiment using a slightly revised design, in which subjects were told that they "owned" the endowed good, made their choice while in possession of the endowed good, and indicated their choice of objects on a form that identified the good with which they were endowed. In this version of the experiment, more "mug owners" chose mugs than did "pen owners," consistent with the endowment effect, but the difference was not statistically significant, according to standard convention ( $p = .18$ ). From this



failure to refute the null hypothesis, PZ concluded that their results were inconsistent with the endowment effect, and that their follow-up experiment validated their interpretation of their original experiment.

One cannot disprove a hypothesis with a null result, of course, especially when the result is directionally consistent with the hypothesis but happens to fall short of statistical significance. PZ are wrong to imply that their follow-up experiment results disprove, or even significantly undermine, the case for the endowment effect's existence. The  $p$  value of .18 reported by PZ means that if, in fact, there were no endowment effect, the chance of finding the difference between the two groups as large or larger than the one PZ report would be approximately one out of six—possible, but not very likely. Given the number of trading experiments that find a statistically significant endowment effect, it seems more likely that the difference between PZ's subjects also indicates an endowment effect rather than random noise, even though their results do not reach the 95% certainty threshold that social science convention requires for the pronouncement of a statistically significant difference between two groups.

Knetsch and Wong (2009) (KW) conducted a subsequent set of experiments that further undermines PZ's criticism of the original trading experiments. In the first test, KW showed subjects a mug or a pen but did not give them physical possession of either. They then told the subjects that, as the result of random assignment, they owned either the mug or the pen, and that other subjects were given the opposite item. Finally, subjects were given a decision form on which they indicated which item they wished to take home with them. As in PZ's initial trading experiment, KW found no endowment effect. In subsequent tests, however, KW produced an endowment effect by making slight changes to the protocol: first by asking subjects to decide between "keeping" their item and "trading" for the other (rather than asking which item they wished to take home), and second by giving subjects physical possession of their endowed item. These results suggest a different interpretation of PZ's results: rather than merely being an artifact of poor laboratory practices, the existence and size of the endowment effect likely depend on contextual features that can make the fact of endowment status more or less clear and/or salient to decision makers.

## 1.6 Concluding Thoughts

It is important to remember that, unlike the Coase Theorem, the endowment effect is not a theoretical prediction, but rather an empirical finding. As such, although the effect has proven robust across a range of contexts, there is no a priori reason to believe that the effect will be equally pronounced, or even exist at all, in all contexts. The evidence suggests that an individual is likely to value a particular entitlement more if he is endowed with it than if he is not, but this does not mean that the facts that trigger endowment status are either clear or unchanging, or that endowment status will have a significant effect on valuation regardless of what other contextual features are present.

Because it would be impossible to test for the presence of the endowment effect in every context in which legal policymakers might have an interest, predicting the presence of the endowment effect in an unexplored context based on its existence in a similar context is unavoidable. The alternative, assuming no endowment effect exists short of a definitive demonstration in the precise context of interest, is impractical in the applied discipline of law, in which decision-makers must act in one way or another in the face of real-world problems. However, legal scholars need to take care to ensure the closeness of their contextual analogies and not to lose sight of the fact that their conclusions will often be contingent on the soundness of such analogies.

## 2 CAUSES OF THE ENDOWMENT EFFECT

---

The endowment effect is important for legal scholars because it suggests that endowment status will affect preferences and behaviors of individuals subject to the law. Specifically, when the endowment effect operates, we will expect to see fewer reallocations of entitlements through private ordering than the Coase Theorem suggests. The primary normative questions that follow for law are whether and how the endowment effect should be taken into account when assigning or reassigning entitlements, facilitating private ordering, or using the legal system to protect entitlements. Addressing these questions appropriately requires an understanding of the endowment effect's underlying cause or causes.

### 2.1 Wealth Effects and Liquidity Constraints

As an initial matter, it seems safe to exclude wealth effects as plausible general explanatory accounts of the phenomenon. An unendowed rich man can be expected to be willing to pay a higher dollar amount to obtain an entitlement than a poor man, all other things being equal, but the rich man's wealth usually would be expected to have the same effect on the amount he would demand to part with the entitlement if he is endowed. Thus, although the rich will often value particular entitlements more than the poor, they usually should be no more likely to exhibit an endowment effect in the general case.

In theory, it is conceivable that because an actor endowed with the entitlement at issue is marginally richer than the same actor who is not endowed, and because richer people are likely to place a higher monetary value on entitlements, the wealth represented by the endowment, not endowment status per se, could generate a higher valuation. This explanation seems implausible, however, when the endowment at issue is a very low-priced item and yet WTA values significantly exceed WTP values. In other words, increasing an actor's total wealth by the quantum of one \$6 mug is not likely to

suddenly make him willing to spend twice as much money on consumer goods going forward.

In addition, many of the endowment effect experiments explicitly control for wealth effects by providing endowed subjects with the entitlement at issue and unendowed subjects with a cash equivalent, thus holding the wealth of subjects constant across experimental groups. Other experiments control for wealth effects by comparing sellers (endowed), not only with buyers (unendowed), but also with “choosers” who may choose between the entitlement and cash. Choosers are as wealthy as sellers (because they can get the entitlement for free), but they behave like buyers, suggesting there is more to the buyer-seller discrepancy than a wealth effect (KKT 1990).

Liquidity constraints theoretically could produce an endowment effect where particularly expensive entitlements are at issue. A person of modest means might report a higher WTA value than WTP value for life-saving surgery, for example, if she lacks the cash and borrowing ability to bid as much for the operation as she would demand if she were entitled to it. Again, however, most endowment effect experiments involve entitlements of small value compared to the wealth of subjects, and in many the experimenters give subjects enough cash to purchase the relevant entitlements. Thus, although liquidity constraints could cause an endowment effect in particular cases, endowment status itself seems to be driving force behind the effect across the variety of experimental studies.

## 2.2 Loss Aversion and Related Explanations

Why would endowment status affect valuation? The endowment effect is routinely explained as a consequence of “loss aversion,” the observation that losses are experienced as more significant than equivalent gains (Thaler 1980; Kahneman, Knetsch, and Thaler 1991; Tversky and Kahneman 1991). To economists, opportunity costs have equivalent value to out-of-pocket costs. But the evidence of loss aversion demonstrates that the two do not feel equivalent, at least in some circumstances. Having a dollar taken away hurts more than not receiving a dollar expected. Loss aversion suggests that gaining relative to an applicable reference point will be perceived as less significant than losing relative to that reference point. If this is so, it follows that individuals will be willing to pay less to *gain* the entitlement (WTP) than they will demand to *lose* the entitlement (WTA), even in the absence of transaction costs.

Although the concept of loss aversion is a useful way to explain the endowment effect, it does not actually help to resolve the puzzle of *why* people appear to value losses more than equivalent gains. That is, loss aversion is an empirical finding rather than a psychological explanation. Some commentators construe the loss aversion explanation for the endowment effect narrowly and apply it only to the specific hypothesis that ownership causes people to place a higher value on the *substantive* entitlement at issue (Klass and Zeiler 2013). But it is also possible that losses are experienced as more significant than gains, and loss aversion results, if endowment status affects how people

view the act of engaging in a transaction or the likelihood of future regret, or even the manner in which people evaluate the expected gains and losses of transacting. All of these psychological explanations for loss aversion, and thus the endowment effect, are inconsistent with the predications of the Coase Theorem. But each one potentially has different implications for normative legal policy designed to respond to the endowment effect, so it is important for legal theorists to attempt to distinguish between them. Unfortunately, the available evidence suggests that each of these explanations is plausible in some circumstances, but none can account satisfactorily for every experimental demonstration of the endowment effect. The best conjecture, at present, is that at least several of these psychological explanations probably play some role in causing the endowment effect, in at least some contexts.

### 2.2.1 *Attachment to Substantive Endowments*

If a tangible entitlement becomes imbued with meaning external to the entitlement itself, it would be incorrect to conclude that a difference between WTA for that imbued item and WTP for an unimbued item demonstrates an endowment effect. For example, I would demand far more money to sell my deceased grandfather's prized pocket watch (which I now own) than I would pay to buy a physically identical pocket watch *that did not belong to my grandfather*. But this compares an heirloom to a commodity. The endowment effect requires that a difference in valuation be based solely on the location of an entitlement. For there to be evidence of an endowment effect, my WTA for my grandfather's watch would need to be higher than my WTP for my grandfather's *actual* watch, assuming a third party had come to possess it.

But the endowment effect could result from an attachment to substantive entitlements that forms solely as a consequence of ownership or possession. Once *a* widget becomes *my* widget, perhaps I like the widget more (Kelman 1979). This explanation for the endowment effect seems implausible when the entitlement in question is held solely for its ultimate exchange value, rather than for personal use, as is the case with the many endowment effect experiments using lottery tickets. A lottery winner would presumably suffer no ill effects from relinquishing the ticket in order to obtain her winnings, so it seems unlikely that trading the ticket prior to the lottery would create such an effect. Attachment has more potential as a psychological explanation for the endowment effect when experimental subjects endowed with tangible property (such as mugs) that they may take home and consume value those items more than the subjects sitting next to them who were not endowed, but even here it seems difficult to explain why endowment alone would create instant attachment observable as a gap between WTA and WTP values.

Attachment becomes a far more appealing explanation for the endowment effect when viewed as a general heuristic, most likely unconscious, that promotes self-protection on average, rather than a logical explanation of why particular entitlements are more valuable when owned than unowned. Evolutionary theorists hypothesize that there was a survival advantage in the evolutionary environment to expending more effort protecting one's resources than in taking away resources from a competitor

(Stake 2004; Jones 2001). Psychologists argue that a way that the mind copes with the fact that we can't always get what we want might be to like what we have more than what we don't have, all other things being equal (Weaver 2008). Either or both explanations might explain an instantly occurring attachment effect, or even potentially an attachment effect that takes root when an individual believes that an entitlement is likely to come into his possession soon.

Recent experiments demonstrating that nonhuman primates exhibit an endowment effect support the explanation of the endowment effect as deeply ingrained in our evolutionary heritage. Lakshminaryanan, Chen, and Santos (2008) gave capuchin monkeys tokens with which they could purchase fruit discs or cereal chunks from the experimenters, and found that the monkeys spent exactly half their tokens on each treat. When endowed with either fruit discs or cereal chunks and allowed to trade for the other treat, however, the monkeys overall—and each subject individually—chose to exchange only a very small percentage of the endowed treats for the other. The effect held even when the experimenters offered the subjects an additional incentive of oats to account for any perceived transaction costs.

Brosnan et al. (2007) found that 58% of their chimpanzee population preferred a peanut butter snack to a frozen juice bar when given the choice between the two food items, but 79% who were endowed with the peanut butter and offered to trade for a juice bar kept the peanut butter and 58% endowed with a juice bar kept it rather than trading for peanut butter (see also Jones and Brosnan (2008)). The chimps willingly traded toys back and forth with the experimenters, so there was no indication that they were unwilling to engage with humans, viewed transaction costs as particularly high, or feared being tricked and left with nothing. When food was at stake, however, they demonstrated an enhanced preference for the item they possessed.

### 2.2.2 *Transactional Disutility*

Rather than reflecting differential valuations of the underlying substantive entitlement at issue based on endowment status, the endowment effect could be caused by differential responses the endowed and the unendowed have to engaging in particular transactions, even when transaction costs, as usually understood, are very low.

Although market transactions are a fixture of modern societies, there might still be some items that people will resist evaluating in terms of dollars, causing sellers to demand a high price for an entitlement to compensate them for the inherent value of the entitlement and the discomfort of subjecting the entitlement to a market transaction, and causing buyers to offer prices lower than an entitlement's inherent value to compensate for the discomfort associated with participating in the transaction (Kelman 1998). For example, Abel might despise neighbor Baker's loud stereo but feel so uncomfortable with the thought of paying Baker to turn the volume down that Abel's WTP is very low. But if the city government were to classify loud music as a nuisance, Abel might feel so uncomfortable with the idea of accepting money from a neighbor that his WTA to not enforce the law might be extremely high. A closely related idea is that people have a desire to "close transactions" (Kelman 1979) and they often prefer not

to continually update their valuations of various entitlements. Willingness to sell only if the price is high enough to be clearly worthwhile, or willingness to buy only if the price is low enough to be clearly worthwhile, could reflect this intuition.

The endowment effect could be driven by transactional disutility on the part of buyers and sellers, but it is also consistent with the intuition that participating in some transactions will create disutility for sellers but not for buyers. The ordinary consumer goods often used in endowment effect experiments are commonly purchased by ordinary individuals, suggesting that WTP values are unlikely to be depressed by a distaste for commodification. But most individuals do not sell consumer goods unless they are engaged professionally in the business of selling such items. It is possible that the distaste associated with selling even common items might cause individuals who are not frequent sellers to demand compensation for that discomfort in addition to the inherent value of the goods. This intuition is consistent with the finding that the endowment effect is less likely when subjects are frequent market participants.

In some contexts, legitimacy norms proscribe the sale of some entitlements, such as those that explicitly involve health and safety, personal integrity, and environmental protections, although there often is not an equally strong normative injunction to buy such entitlements if they are unowned. In one experiment, Boyce and colleagues (1992) found that a disparity between subjects' WTA and WTP for a small tree increased dramatically when subjects were told that any trees left in the possession of the experimenters would be destroyed. This additional piece of information caused the subjects' WTA values to increase substantially while WTP values remained constant, suggesting that subjects felt it would be wrong to sell a tree that would then be destroyed, but that their moral obligation did not extend to paying to protect a tree. This might be because a single seller can seem uniquely morally responsible, whereas moral culpability for not buying is diluted because there are many good causes in which to invest and also many other nonbuyers with whom to share culpability for not buying.

### 2.2.3 *Regret Avoidance*

Another potential explanation of the endowment effect, rooted in the participation in transactions rather than the valuation of the underlying substantive entitlement itself, is that valuation uncertainty leads to risk-averse trading behavior in order to minimize the risk of ex post regret. When an individual engages in any transaction in which his preferences are at all uncertain, there is a risk that, with hindsight, he will decide that the transaction actually reduced his total utility, which can cause him to regret his choice. Owners might be willing to relinquish an entitlement only if offered the inherent value of the entitlement plus a premium to compensate for the possibility of future regret, and nonowners might be willing to pay only the expected value of the entitlement less a discount to compensate for the risk of future regret. Differences in WTP and WTA prices thus could be seen as a regret-avoidance premium (Loomes and Sugden 1982).

The regret-avoidance explanation can explain the endowment effect, however, only if the utility consequence of regretting a "bad" decision to *engage* in a transaction is



greater than the utility consequence of regretting a “bad” decision *not to engage* in a transaction, and the utility consequence of *regretting* making a “bad” decision is greater than the utility consequence of *rejoicing* over making a “good” decision. There is reason to suspect that both of these assumptions are correct, at least often.

As to the former, it is plausible to think that actions are more salient than failures to act, and thus have the potential to cause greater regret. After all, in the course of a day, we make relatively few decisions to act affirmatively (i.e., engage in transactions) compared to all the decisions we implicitly make not to act affirmatively (i.e., not engage in possible transactions). Empirical research is consistent with this intuition about regret (Gilovich and Medvec 1995) and supports the claim that people tend to prefer harmful omissions to harmful commissions (Ritov and Baron 1990).

The latter assumption is less intuitive but finds support in empirical research as well: although people are happy when they can attribute decisions that look good in hindsight to their own agency, that positive emotion appears to be weaker than the unhappiness that results when their choices lead to decisions that look bad in hindsight (Taylor 1991).

The regret-avoidance theory is consistent with the observation that the disparity between WTP and WTA increases as the value of the entitlement becomes more uncertain, and with the somewhat peculiar specific finding that there is no endowment effect for tokens with a fixed cash value while there is an endowment effect for tokens with an uncertain cash value within a specified range (van Dijk and van Knippenberg 1998). There is little likelihood of future regret if one were to sell a token worth \$5 for exactly \$5. There is a possibility of future regret, however, if one were to sell a lottery ticket for \$5 at it later pays off \$10, even if the *ex ante* expected value of ticket was \$5.

Two recent studies lend further support to this theory. Weaver and Frederick 2012 (WF) show that when subjects have information concerning the market price of the entitlement at issue, buyers bid their use valuation of the entitlement (which is lower than market price, on average, because most people value most goods below market price), but seller bids are anchored by the market price. These results are consistent with findings by Simonson and Drolet (2004) that when subject predictions of typical store prices of items are manipulated with anchor values, WTA valuations are affected more than WTP valuations. WF infer that greater seller focus on external prices indicates that endowed subjects worry about getting an objectively bad deal, whereas unendowed subjects are not symmetrically concerned with missing out on an objectively good deal if the price exceeds their subjective value.

Arlen and Tontrup (2013) (AT) endowed subjects with a lottery ticket labeled “heads” or “tails” redeemable for 8 euros if it matched the result of a forthcoming coin toss. Subjects in a baseline group were offered the opportunity to trade their tickets for tickets with the opposite label plus 25 cents. Subjects in a second group were told that an agent subject (with a financial incentive to recommend trading) would decide for them whether to trade, but that they then could veto the agent’s decision if they wished. Seventy percent of subjects in the baseline group refused to trade, but 69% of subjects in the agency group declined to exercise their veto over the agent’s decision to trade. Given that subjects knew both that their agents had a profit incentive to recommend



the trade and had no private information about the potential outcome of the coin flip, AT contend that the agent's recommendation could have no plausible effect on the subject's decision other than to lessen the subject's responsibility for the decision, and thus protect against ex post regret that she might feel if she were to lose the coin flip.

#### 2.2.4 *Attention*

Johnson, Haubl, and Keinan 2007 (JHK) propose that the endowment effect is best explained by query theory, a more general psychological theory that preferences depend on the order in which an actor considers costs and benefits of a potential change in status. Specifically, JHK contend that endowment status affects the attention paid to the different consequence of transactions. They propose that, in the face of a potential transaction, people first consider the advantages of the status quo and then the advantages of the alternative. A primacy effect in memory causes us to better recall the results of the first inquiry, which results in both buyers and sellers placing greater mental emphasis on the benefits of the status quo relative to the alternative. The consequence is a bias in our prediction about the utility consequences of trading or not trading. In other words, the aversion to losses comes from the effect that the reference point has on the cognitive process individuals often use when evaluating the costs and benefits of transacting.

JHK provide support for this theory by conducting a familiar mug experiment but asking both "sellers" (endowed with the mug) and "choosers" (not endowed) to write down all the reasons they would rather have either the mug or the money prior to determining their valuations. Sellers, on average, valued the mug at a significantly higher price than choosers (\$5.71 to \$3.42), with sellers identifying significantly more positive aspects of keeping the mug than of trading for money, and choosers identifying more positive aspects of having money than of obtaining a mug. The experimenters then directed sellers to first list the positive aspects of trading mug for money (followed by negative aspects) and choosers to first list the positive aspects of obtaining the mug rather than money (followed by negative aspects). With this manipulation, the endowment effect disappeared almost entirely

## 2.3 Concluding Thoughts

The weight of the evidence suggests that endowment status affects valuation, on average, in many contexts, by acting as a reference point that distinguishes between perceived "gains" and perceived "losses." But attributing the endowment effect to loss aversion does not, in itself, identify a psychological mechanism. The endowment effect could result from attachment to endowments, disutility associated with transacting, regret aversion, and/or differential attention paid to the status quo compared to alternatives.

To make matters more complicated, it seems likely that the endowment effect is caused by a combination of these factors, not just a single one (Sayman and

Onculer 2005). In a study using bottles of wine, for example, Georgantzis and Navarro-Martinez (2010) found that subjects' higher WTA prices (compared to WTP prices) were attributable to both greater positive feelings for the wine resulting from endowment *and* uncertainty concerning its value, suggesting that both attachment and regret avoidance were at play. Regression analysis conducted by JHK (2007) indicates that their attention hypothesis explains only a portion of the endowment effect that they measured. Furthermore, although instructing sellers and choosers to list aspects of the proposed transaction in the opposite order of what seems most natural succeeded in eliminating the endowment effect, it did not create a reverse, "nonendowment effect," as would be expected if aspect ordering were the endowment effect's sole cause.

The difficulties posed by the lack of clarity concerning the endowment effect's precise causes for scholars who wish to evaluate the effect's normative implications are considered in Section 3.

### 3 THE ENDOWMENT EFFECT AND ECONOMIC ANALYSIS OF LAW

---

Economic analysis often adopts the normative position that entitlements should be allocated efficiently, usually understood to mean that they should be allocated to the claimants likely to place the highest monetary value on them. By demonstrating that claimants will often lack a single valuation for an entitlement, evidence of the endowment effect raises substantial conceptual challenges to this analysis. Legal scholars have risen to the challenge by exploring the impact of the endowment effect in a wide variety of areas. Space limitations prevent this chapter from considering more than a fraction of such attempts. At the same time, as this section attempts to demonstrate, legal scholarship has yet to reach consensus concerning how the endowment effect should affect policy analysis.

#### 3.1 Assigning Entitlements

Law and economics analysis typically assumes legal entitlements should be assigned to the claimants who value them the most. Any other assignment will create transaction costs, as higher-value claimants purchase the entitlements from lower-value claimants. If transaction costs are too high, the result could be allocative inefficiency, as higher-value claimants fail to obtain them from lower-value claimants (Posner 2007, §3.6). The endowment effect heightens the case for "market mimicking" allocations by indicating that that initial assignments will be "sticky" even if the transaction costs associated with reallocation are low. But the endowment effect also raises conceptual

problems for market mimicking because it is conceivable that some parties could have higher WTP values while others have higher WTA values.

Some scholars have proposed that one of the two value measures, WTA or WTP, always be respected, but such approaches tend to implicitly assume a particular cause of the endowment effect. Herbert Hovenkamp, for example, contends that wealth is maximized when entitlements are allocated to the claimant with the highest WTA value (Hovenkamp 1990, 1991, 1994), but this depends on his assumption that the cause of the endowment effect is that some individuals are wealth and liquidity constrained and simply cannot afford to bid as much for entitlements as they actually desire them. Hovenkamp's argument is defensible given that assumption, but not if other factors drive the endowment effect.

I have argued that the theoretically proper entitlement allocation demands close attention to situation-specific causes of the endowment effect. If the differential between WTP and WTA values is explained by attachment, WTA values should serve as the basis for allocation decisions because this will ensure entitlements are granted to those who will value them most *ex post*. But if the differential is explained by transactional disutility that accompanies selling but not buying, then WTP is a more accurate measure of the claimant's value for the entitlement at issue. This is because if the state does not initially award the entitlement to the claimant, the claimant will never actually suffer the disutility or fear of *ex post* regret that he would suffer if he were to voluntarily give up the entitlement (Korobkin 1994). This approach would run into severe practical obstacles, of course.

Cass Sunstein (1986) has argued that the endowment effect demonstrates market mimicking is an incoherent allocation strategy and ought to be abandoned. He concludes that the endowment effect supports "considerable legislative and judicial intrusion into private preference structures," rather than basing legal structures on existing private preferences that are clearly context-dependent. Because law can shape private preferences, Sunstein argues that the law should attempt to do so for the better, as defined by policymakers, rather than react to existing revealed preferences in the name of efficiency, as traditional economic analysis would do.

One way to avoid the problem of choosing between WTP and WTA measures without giving up on the concept of efficiency in resource allocation might be for the law to attempt to obscure the reference point that an endowment creates, in hopes of minimizing any endowment effect in an effort to spur private ordering. This approach might favor context-specific legal "standards" over bright-line "rules" (Korobkin 2000). A factory owner might have a very high WTA for installing pollution mitigation equipment if he has a clear right to pollute, and the factory's neighbors might have a very high WTA for installing air conditioning and HEPA filters on their houses if they have a clear right to enjoin the factory's operation. But if the law permits only "reasonable" amounts of air pollution, clouding who is legally entitled in the particular case, the parties might be more willing to determine which can mitigate the harm more cheaply and jointly fund that technology. This theory finds some support in the well-known fact that most lawsuits are settled prior to adjudication, when endowments are in dispute,

but that litigants rarely renegotiate entitlements after a court has finally adjudicated a case and thus clarified which party owns the disputed entitlement (Farnsworth 1999).

Similarly, the law might attempt to create multiple or competing reference points, obscuring which party has a possessory interest or expectation of a particular course of action. Ginsburg, Masur, and McAdams (forthcoming) suggest temporary legislation that reverses entitlements after a set amount of time. The policy goal of these authors is to generate better information on the costs and benefits of particular legal rules. For example, they suggest a law prohibiting smoking in bars that sunsets after a set amount of time will produce data concerning the costs and benefits of laws that either allow or prohibit smoking. But shifting entitlements could also potentially facilitate private ordering (in this case, some bars deciding to allow smoking and others deciding to prohibit it) by reducing the attachment that either group of claimants feel toward any particular equilibrium. During the 2012 presidential campaign and the “fiscal cliff” negotiations that followed, Republicans attempted to frame changes from Bush-era tax rates as tax increases, while Democrats argued that the appropriate reference point for discussion was the higher rates that would apply by law when those “temporary” rates expired. Perhaps the uncertainty as to the appropriate reference point undermined the bias toward policy conservatism that the endowment effect can create, enabling some congressmen to vote in favor of the successful tax bill that increased rates for some taxpayers relative to the rates in place on December 31, 2012, but not relative to the tax rates that took effect one minute past midnight on January 1, 2013 (Public Law 112-240, 126 Stat. 2313).

Should the policy response to an attachment-caused endowment effect be different if the entitlement in question is nonrivalrous—that is, it can be both enjoyed and transferred? Buccafusco and Sprigman (2011) (BS) argue that it should. They claim that the fact that creators of intellectual property that can be enjoyed by one individual and simultaneously licensed to others exhibit higher WTA prices than noncreators (in an experiment that they conducted) counsels for vesting rights in sponsors of creations, rather than in creators themselves, in order to facilitate more licensing transactions. The key to their argument is that, when enjoyment of an entitlement is nonrivalrous, assigning licensing rights to parties with lower WTA will not have the attendant cost of depriving high-WTA creators of their enjoyment value, as is the case when consumption is rivalrous. Even if painters adore their paintings more than anyone else, the argument goes, we don’t maximize social welfare by giving painters the right to sell copies (which they will do only at high prices), because painters can still spend all day staring at and enjoying their works even if someone else has the right to sell copies.

BS assume, but do not actually demonstrate, that their “creator” subjects have greater endowment effects than noncreators. Although the creators demonstrated higher WTA prices than noncreators, they might have had equally stratospheric WTP prices—BS did not study the latter measure for the creator group. If creators have WTP prices that exceed the WTA prices of noncreators, assigning rights to exploit the rents of the creation to noncreators would serve only to generate transaction costs, as

creators would purchase the rights in question from the noncreator entitlement holders (absent liquidity constraints).

But even assuming that creators have a pronounced endowment effect—that is, abnormally high WTA prices but ordinary WTP prices—there is a substantive weakness to the normative conclusion suggested by BS. Even though consumption of some intellectual property is nonrivalrous, such that creators could continue to enjoy creations that are simultaneously licensed to others, the rents available from licensing are limited. Any proposal that involves depriving creators of property rights must take into account the dynamic effects on production that such a policy will certainly have. Even if vesting such rights in noncreators would facilitate transactions in nonrivalrous goods once those goods exist, it would reduce the incentive to create, potentially leading to an inefficient undersupply of creations.

### 3.2 Redistributing Rights

The government—in the form of courts, legislatures, and administrative agencies—often faces the issue of whether it should redistribute an entitlement that is allocated *ex ante* to one class of claimants. The endowment effect leads to the prediction that, once established, altering the status quo will be difficult because parties who benefit from the status quo will fight harder to avoid a change in the status quo than they would have fought to establish their preferred position in a counterfactual world. This prediction is difficult to test because it mirrors the prediction of public choice theory that the regulatory status quo is difficult to change due to the entrenchment of interest groups. At a minimum, however, the endowment effect is likely to increase the inertia in favor of existing regulatory policies. More speculatively and interestingly, the endowment effect might well provide a better primary explanation for regulatory inertia than does public choice theory because it can be argued that, in the absence of the endowment effect, rational interest groups should be expected to exert as much effort in the political process to secure potential gains as to prevent potential losses of the same magnitude (Zamir 2012).

A corollary to the prediction that the endowment effect will entrench the regulatory status quo is that imposing new or more stringent regulations on existing entitlements will tend to be disfavored relative to regulating new entitlements (Sunstein 1998). All other things being equal, regulating new entitlements is likely to generate less opposition because the entitlement holders will lack a sense of endowment in the regulatory status quo. For example, imposing new pollution standards on existing sources of greenhouse gases in order to mitigate global climate change will be difficult because losses, in the form of economic dislocation, will loom larger than the potential environmental gains (Rachlinski 2000). On the other hand, regulating future sources of greenhouse gases may be a more politically promising strategy.

Is a bias of law in favor of the status quo normatively appropriate? Without the endowment effect, an efficiency analysis of a proposed redistribution of an existing

good or right would be simple, at least in theory: compare the value placed on for the entitlement of the would-be beneficiaries (“winners”) to the value of the entitlement to the would-be burdened class (“losers”). When endowment status affects valuation, however, this comparison would tend to lead to conservative policy choices with a strong bias in favor of the status quo, because the winners’ WTP values (which will tend to be lower) would be compared to the losers’ WTA values (which would tend to be higher). The bias clearly has political implications, as it suggests that society’s “haves” will prevail over its “have-nots” in policy disputes even more often than under traditional economic analysis, which itself is often criticized for favoring those who have the dollars to back up their preferences relative to those who do not.

If the high WTA value of the burdened group reflects that it would experience greater negative utility from redistribution than the remainder of the polity would experience positive utility, as, for example, the attachment explanation would suggest, redistribution can be viewed as inefficient and, thus, normatively undesirable. The implicit value judgment inherent in this proposition is that the effect that entitlement status has on the strength of individuals’ preferences is as relevant to individual utility, and hence to social welfare, as an entitlement’s tangible attributes are. If I value a mug in part because of the use I can put it to, in part because of the enjoyment I will get from looking at it, and in part because I am attached to it (or not), the third factor that contributes to my overall utility is no less deserving of respect than the first two.

But this conclusion is, at least, contestable. We humans are noticeably bad at affective forecasting, and we usually adapt quickly to changes in circumstances (Korobkin 2011). If I suffer negative utility when I lose an entitlement in part because of the attachment I feel toward it, perhaps I will soon become just as attached to the entitlements I receive in compensation (or purchase with the dollars I receive in compensation) and end up no worse off. This seems especially likely if the attachment I feel to my endowments is the result of an evolutionary instinct I have to protect what I possess, an instinct likely to produce little survival advantage in the twenty-first century. This conjecture finds support in a recent experiment. Engelmann and Hollard (2010) randomly endowed subjects with one of two consumer goods of approximately equal market value (Round 1). Some were allowed to trade with a subject endowed with the other good if they wanted to, while others were required to trade. All subjects were then endowed with one or two other goods as compensation for participating in a survey and allowed to trade with another subject if they wanted (Round 2). Subjects who were forced to trade in Round 1 demonstrated no endowment effect in Round 2—perhaps because they learned how easily they adapted to changed circumstances—but subjects not forced to trade in Round 1 did demonstrate an endowment effect in Round 2.

Where the endowment effect is driven by transactional disutility or regret avoidance—especially if caution toward transacting is experienced more by sellers than buyers—a strong argument can be made for ignoring WTA values and redistributing entitlements to individuals or groups with higher WTP even when the winners’ WTP would not exceed the losers’ WTA (and, therefore, voluntary exchange would not occur). The logic is that if a voluntary exchange is impeded by the owner’s desire not

to commodify the entitlement or fear of later regretting a decision to sell, involuntary redistribution would avoid those costs, so they are property ignored in an efficiency analysis (Korobkin 1994, 1998a).

Like the consequences of attachment as an endowment effect explanation, however, this normative account of how legal policy should respond if the endowment effect is driven by transaction disutility is contestable, for at least two reasons. First, the claim that transaction disutility should be ignored in theory if it can be avoided in practice implicitly assumes that decisional autonomy itself has zero value (Korobkin 2011; Blumenthal 2007). The freedom to make one's own decisions certainly has utility consequences. If so, policymakers interested in efficiency would need to balance the costs to autonomy of forced redistribution against the potential benefits that would result from increased direct enjoyment of entitlement ownership. It is quite possible that the loss of autonomy suffered by the losers of an entitlement reallocation would create greater disutility than would be associated with voluntarily transacting (Lewinsohn-Zamir 2012).

Second, forced distribution—and both the cost to autonomy and the potential for error when government actors incorrectly believe that the winners value the entitlement more than the losers do—is arguably unnecessary if entitlement holders have a way to avoid the potential transactional disutility associated with selling. Recall that AT demonstrated experimentally that a majority of experimental subjects without an agent declined to trade a lottery ticket for a ticket with identical odds of winning plus 25 cents, but a majority of subjects assigned an agent who recommended the trade (as the agents' publicly transparent compensation structure incentivized them to do) declined to veto the agent's recommendation (Arlen and Tontrup 2013). In another variation of the experiment, AT also found that half of principals given the choice of making the decision themselves or delegating it initially to an agent chose to delegate, and nearly all of those who delegated subsequently abided by the agent's decision rather than vetoing it, as they were permitted to do. That is, at least in a context in which an endowment effect results from the fear of regret associated with selling an entitlement of uncertain value, subjects seemed able to reduce transactional disutility enough to enter in a transaction with positive expected value by delegating the actual decision.

### 3.3 Facilitating Private Transactions

In addition to allocating or reallocating legal entitlements, law facilitates the reallocation of clearly established entitlements between private parties. The endowment effect has consequences for the use of law for this purpose as well.

Most contract law rules are “default” rules rather than “immutable” rules (Ayres and Gertner 1989). That is, the rules apply to parties that fail to negotiate their own terms to govern their transaction, but parties may opt for different rules by mutual agreement. The content of a default term is not a legal endowment in the sense that a property right is. Unlike mugs, lottery tickets, or even public goods, which an entitlement holder can use and enjoy without the assent or assistance of anyone else, a party that stands to



benefit from a default rule can do so only if she is able to convince another party to contract with her and accept that term (Zamir 1997; Korobkin 1998a). Still, since default terms will govern the parties if a contract is reached unless another term is chosen, it seems plausible that they might serve as reference points from which actors judge changes as either “gains” or “losses,” in the same way that property rights or physical possession often serve as such reference points.

In a series of experiments I conducted, law students were asked to play the role of a lawyer negotiating a shipping contract on behalf of an overnight delivery company (Korobkin 1998b). In one type of scenario, subjects were told that there were two possible terms that could be used in the contract to deal with a particular issue, one of which clearly favored the subjects’ client and the other that clearly favored the counterparty. Half of the subjects were told that the term favoring their client was the default and were thus asked to provide their WTA price (on a per-package basis) to agree to insert term *B* into the contract, thus contracting around the default term. The other half of the subjects were told that term *B* was the default term and were asked to provide their WTP for inserting term *A* into the contract, thus contracting around the default. In both scenarios, WTA values significantly exceeded WTP values; in other words, the strength of the subjects’ preferences for the term that benefited their client was biased in favor of whichever term was identified as the default.

To be sure, contracting parties might demonstrate a preference for default terms for reasons other than the endowment effect. For example, individuals without complete information about the costs and benefits of various terms might view the state’s choice of a default term as signaling information about the quality of that term that increases the term’s desirability. Parties might assume that most other sets of parties will accept the default term and that using the most popular term would therefore provide network benefits. Or parties might prefer a default term out of fear that suggesting an alternative might signal to the other party that the party seeking to avoid the default is likely to be an undesirable contracting partner (Ben-Shahar and Pottow 2006). But default terms seem to have a stronger effect than can be explained by any of these hypotheses. My experiments demonstrate that the bias in favor of default terms can persist even when the economic value of that term to the parties is clear (thus controlling for the signaling value of a term), when contracting around the default does not require the party to propose doing so (thus controlling for the possibility of signaling one’s undesirability as a trading partner), and even when parties are told that what is perceived as the default term will be accepted by only a minority of parties (thus controlling for the network externalities explanation).

These experiments demonstrate the positive point that the selection of default rules can have a substantive effect on private ordering. As usual, the normative implications are less obvious. How should legal policymakers behave differently knowing that the endowment effect exists for contract terms?

I have argued that the evidence suggests three basic policy prescriptions. First, because the endowment effect makes default terms sticky and reduces the number of parties expected to contract around defaults, it is particularly important to select

default terms that maximize efficiency for most contracting parties. Second, because some parties will not contract around default terms that are objectively less efficient than alternative terms, “tailored” defaults (defaults based on the factual circumstances of individual transactions) are preferable to “untailored” defaults (defaults that apply to all contracting parties). Third, for contingencies that are highly salient to parties such that they are unlikely to forget to negotiate terms to address them, it might be preferable not to provide a default term at all, instead denying enforcement of contracts that fail to provide a term to govern that contingency (Korobkin 1998b). This analysis assumes that policymakers care only about maximizing the utility of the contracting parties and have no substantive preference for any specific contract compared to any other. If a contract term will create externalities, then the stickiness caused by the endowment effect can be employed to encourage that more socially beneficial term without mandating it (Korobkin 2009).

It has also been suggested that the stickiness of default terms suggests that policymakers can achieve redistributive aims through their choice of contract default rules. To take one example, Millon (1998) contends that establishing a “just cause” default term for termination in employment contracts (rather than an “at will” default) would transfer wealth from employers to employees. The reasoning of the endowment effect suggests that employers who bargain around a just-cause default for an at-will term would have to offer employees higher compensation, making employees better off than they would be under an at-will default, because employees who feel endowed with just-cause protection will value it more highly than those who feel unendowed. But the increased employee preference for just-cause protection that results from the quasi-endowment will also cause some employers to accept a just-cause term and offset its cost in other aspects of the employment contract. Whether these employees would be better off with that trade-off depends on whether their WTA price for just-cause protection accurately mirrors the true value to them of that protection or overstates it—the endowment effect’s foundational riddle.

This implication of the endowment effect for the analysis of contract default rules is complicated by the fact that, as is the case for property rights, a rule of contract law is not the only potential reference point that individuals will use when evaluating whether an alternative course of action would constitute a gain or loss. Specifically, common or standard contract terms can create a reference point, from which parties determine whether they are gaining or losing. Just as the endowment effect can make default rules of contract law sticky, it can make customary terms sticky. In one of my experiments, subjects were told that they had agreed with their negotiation counterpart to use a standard industry form as the basis for their negotiation, but that each term would be individually negotiated. They were also told that the form term was the opposite of the legal default rule. That is, if the subjects contracted around the form term, they would reestablish the default term that the standard form circumvented. In this set of experiments, subjects’ preferences were biased in favor of the form term, rather than in favor of the default rule (Korobkin 1998b).

A consequence of this finding is that market participants with superior legal knowledge and business acumen (usually sellers) can potentially reduce resistance on the part of their contracting counterparts (usually buyers) to terms that favor them by creating alternative reference points, such as the text of standard form contracts. If such tactics are normatively undesirable, the law can perhaps prohibit sellers from taking steps likely to shift the reference point. For example, after a political uproar over banks charging consumers huge fees for attempting to withdraw cash from an ATM or using a debit card when doing so would overdraw their accounts, in 2010 the federal government required that banks only be able to provide such overdraft protection when customers affirmatively opted in, rather than making overdraft protection part of the standard bundle (Willis 2013). But for this approach to promote efficiency, lawmakers either need a theory as to why WTP values are a better proxy for preferences than WTA values, or a theory of market failure external to the endowment effect, such as, for example, that buyers are harmed by seller behavior but usually do not realize this because they do not pay attention to most contract terms.

### 3.4 Rules Protecting Entitlements

Finally, the law protects entitlements by resolving disputes. How we understand the positive and normative consequences of this aspect of law is also affected by the endowment effect.

#### 3.4.1 *Burden of Proof*

When two or more claimants dispute which is endowed with a legal right, they often turn to the courts for resolution. Since the legal entitlement at issue is necessarily disputed in such situations, a lawyer might predict that no endowment effect would be present. But experimental evidence suggests that the defendant's physical possession of resources that the plaintiff seeks causes most observers to perceive the defendant as endowed and the plaintiff as unendowed in this situation. Zamir and Ritov (2012) (ZR) have demonstrated that, in this context, legally trained experimental subjects are unwilling to rule for plaintiffs, and thus redistribute resources to them from defendants, unless the weight of evidence substantially favors a plaintiff, even when the legal standard technically requires only a simple preponderance of the evidence.

If both plaintiffs and defendants accept that the defendant enjoys endowed status (see Rachlinski 1996, Zamir and Ritov 2010), this application of the burden of proof can be defended normatively (at least assuming an attachment explanation of the endowment effect) on the ground that an erroneous ruling concerning which party actually enjoys the legal endowment will harm defendants more than plaintiffs. But this argument loses its force if plaintiffs believe they are endowed with the legal entitlement to the redress that they seek, which is probably the case in at least some circumstances (see Korobkin and Guthrie 1994). In addition, as ZR point out, even if judicial bias in favor of defendants in the resolution of disputes over legal entitlements would maximize ex

ante social welfare, such a bias has the clearly undesirable consequence of encouraging claimants to act with reference to possession rather than the rule of law.

### 3.4.2 *Choice of Remedy*

It has long been recognized that the quality of an entitlement depends on whether it can be vindicated by the power of a judicial injunction (“property rules”) or, alternatively, it can be violated so long as the violator pays a court-determined amount of money damages—usually the amount necessary to make the owner “whole” (liability rules) (Calabresi and Melamed 1972). Traditional economic analysis assumes that the choice of remedy does not affect the entitlement holder’s valuation, and reaches normative conclusions as to the appropriate type of remedy by comparing the transaction costs of property rules arising from the fact that nonowners must bargain with owners to obtain entitlements with the risks of judicial error by courts called on to set the price of liability rules (Polinsky 1980).

In two experiments, Rachlinski and Jourden (1998) (RJ) investigated whether the nature of the remedial protection might affect the extent or even the existence of the endowment effect. The authors had subjects play the role of the manager of an environmental preservation trust that either might have to “buy” environmental protection from corporations that threaten to harm the environment or “sell” rights to create harm. They found a significant endowment effect when subjects were told the entitlements in question were protected by property rules, but not when subjects were told rights were protected by a type of liability rules (Rachlinski and Jourden 1998). RJ conclude that their data suggests the relatively weak protection of liability rules undermines the sense of endowment and weakens its role as a reference point for gain/loss attributions.

There is reason to believe, however, that RJ’s results could be particularly sensitive to the precise context of the experiments. This experiment concerned environmental entitlements, which are particularly suited to the transactional disutility explanation for the endowment effect, and the difference in results between the property-rule subjects and the liability-rules subjects were driven by particularly high levels of unwillingness to sell under the injunction remedy. It is possible that the endowment effect observed under the property rule was motivated by subjects’ belief that it is improper to sell an environmental resource that one can protect, and that this belief was undermined in the liability rule scenarios because the law permitted the destruction of the resource for a price. In other words, the results might have been driven by what the choice of remedies says about society’s commitment to the environment rather than by any differences in feelings of ownership that the choice creates. Thus, while the results certainly show something, it is unclear whether they suggest that liability rules will reduce the endowment effect generally, or just in situations in which high WTA prices under property rules reflect a community perception that selling that type of entitlement is immoral—a perception that could conceivably be weakened by protecting the entitlement with only a liability rule.

If the appropriate positive conclusion to be drawn is unclear, the normative implication also seems obscure. RJ contend that enforcement rules that avoid an

endowment effect are preferable to rules that create an endowment effect. But this assumes that the endowment effect is undesirable, a conclusion that, again, seems to depend on the cause of the WTP-WTA differential in particular contexts. If higher WTA values reflect true attachment to the entitlement, a rule that prevents owners from developing such attachment arguably reduces potential social welfare.

### 3.4.3 *Measuring Damages*

When liability rules require courts to award compensatory damages to compensate an entitlement owner, the endowment effect implies that the value of compensatory damages might differ depending whether the question is framed as the amount necessary to cause a victim to voluntarily suffer a harm or as the amount the victim would pay to rescind the harm.

McCaffery, Kahneman, and Spitzer (1995) (MKS) provide data supporting this implication. Experimental subjects placed in the role of jurors were asked to award damages to a plaintiff for the pain and suffering of an injury. Half of the subjects were instructed to determine how much the plaintiff would need to be made “whole,” while the other half were told to determine the amount that the plaintiff would have required to willingly accept the injury. Responses of the second group were twice as high as those of the first, on average.

Again, the normative implications of such a large difference are debatable. MKS contend that the higher, WTA responses would overcompensate victims. This position is at odds with that taken by some scholars who have addressed the question in the context of environmental harms, who have argued that environmental degradation constitutes a loss from the status quo and should thus be compensated based on WTA values of those who suffer harm (Rutherford et al. 1998; Williams 1995). The MKS approach seems defensible if we assume that WTA estimates are high in these types of cases because victims voluntarily trade an entitlement to bodily or environmental integrity for money, which may be aversive and difficult because people do not like to commodify either their health or their environment. There is no need to compensate victims for this quantum of harm when their entitlements are taken without their consent. But if we assume other causal explanations of the endowment effect, this reasoning might not seem as strong.

The choice of whether compensatory damages should be assessed based on the WTA or WTP values of harmed parties also potentially has implications for the substantive law of torts. Under a standard understanding of negligence law, whether an injury-causing activity is negligent depends on whether the benefits of the activity outweigh the expected cost of the risk *ex ante*, taking into account both the likelihood that the activity would cause damages and the extent of those damages (United States v. Carroll Towing Co. 1947). Under this definition, a WTA-based damage measure would render more acts negligent than a WTP-based damage measure.

### 3.5 Concluding Thoughts

The endowment effect suggests that private parties will trade or alter legal entitlements less often than the Coase Theorem predicts. This friction suggests that the role of the law in assigning, redistributing, facilitating private transactions, and protecting entitlements could be more important than traditional legal economists usually assume when the goal is the efficient allocation of resources. But, at the same time, the challenge the endowment effect raises to the established constructs of preference and value has prevented even legal scholars concerned with promoting efficiency from agreeing on how law and legal policy should respond to the evidence of the effect.

## 4 CONCLUSION

---

This chapter seeks to outline the contours of the endowment effect and illustrate the breadth of the effect's relevance to the analysis of law, along with the substantial challenges to normative analysis that it presents. Regardless of subject area, regardless of whether the legal regime in question exists to allocate or reallocate entitlements, facilitate their transfer, or enforce substantive rules, and regardless of whether the analyst is concerned with positive questions (i.e., what consequences a given legal rule will have on the behavior of those subject to it) or normative questions (i.e., what rule or set of rules ought to be enacted), the endowment effect bears on any complete analysis of the issue.

The concerns, problems, and questions raised by the endowment effect are varied and significant enough to serve as a research agenda for legal scholars for many years, as well as a discussion agenda for law school instructors. As empirical research continues to deepen our understanding, both about the contexts in which the endowment effect does and does not operate and the causes that drive the endowment effect in those contexts, endowment effect analysis hopefully will continue to become more precise, more conclusive, and more useful in the design of legal policy.

## ACKNOWLEDGMENTS

---

Jennifer Arlen, Joe Doherty, Jay Koehler, Doron Teichman, Eyal Zamir, and Kathryn Zeiler provided helpful advice and comments. Jordan Jeffrey and Sam Pierce contributed indispensable research assistance. This chapter draws on Korobkin (2003).

## REFERENCES

- Adamowicz, W. L., V. Bhardwaj, and B. Macnab. (1993). Experiments on the Difference Between Willingness to Pay and Willingness to Accept. *Land Economics* 69:416–27.
- Arlen, J., M. L. Spitzer, and E. L. Talley. (2002). Endowment Effects within Corporate Agency Relationships. *Journal of Legal Studies* 31:1–37.
- Arlen, J., and S. Tontrup. (2013). A Process Account of the Endowment Effect: Voluntary Debiasing through Agents and Markets. Available at: <http://ssrn.com/abstract=2263447>.
- Ayres, I., and R. Gertner. (1989). Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules. *Yale Law Journal* 99:87–130.
- Bar-Hillel, M., and E. Neter. (1996). Why Are People Reluctant to Exchange Lottery Tickets? *Journal of Personality and Social Psychology* 70:17–27.
- Ben-Shahar, O., and J. Pottow. (2006). On the Stickiness of Default Rules. *Florida State University Law Review* 33:651–82.
- Blumenthal, J. (2007). Emotional Paternalism. *Florida State University Law Review* 35:12–72.
- Boyce, R. R., T. C. Brown, G. H. McClelland, G. L. Peterson, and W. D. Schulze. (1992). An Experimental Examination of Intrinsic Values as a Source of the WTA-WTP Disparity. *American Economic Review* 82:1366–73.
- Brosnan, S. F., O. D. Jones, M. C. Marenco, A. S. Richardson, S. P. Lambeth, and S. J. Schapiro. (2007). Endowment Effects in Chimpanzees. *Current Biology* 17:1–4.
- Buccafusco, C., and C. Sprigman. (2011). The Creativity Effect. *University of Chicago Law Review* 78:31–52.
- Calabresi, G., and A. D. Melamed. (1972). Property Rules, Liability Rules, and Inalienability: One View of the Cathedral. *Harvard Law Review* 85:1089–128.
- Coase, R. (1960). The Problem of Social Cost. *Journal of Law & Economics* 3:1–44.
- Cohen, D., and J. L. Knetsch. (1992). Judicial Choice and Disparities between Measures of Economic Values. *Osgoode Hall Law Journal* 30:737–70.
- Coursey, D. L., J. L. Hovis, and W. D. Schulze. (1987). The Disparity between Willingness to Accept and Willingness to Pay Measures of Value. *Quarterly Journal of Economics* 102:679–90.
- Dubourg, W. R., M. W. Jones-Lee, and G. Loomes. (1994). Imprecise Preferences and the WTP-WTA Disparity. *Journal of Risk and Uncertainty* 9:115–33.
- Engelmann, D., and G. Hollard. (2010). Reconsidering the Effect of Market Experience on the “Endowment Effect.” *Econometrica* 78: 2005–19.
- Ericson, K. M. M., and A. Fuster. (2011). Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments. *Quarterly Journal of Economics* 126:1879–1907.
- Farnsworth, W. (1999). *Do Parties to Nuisance Cases Bargain after Judgment? A Glimpse inside the Cathedral*. *University of Chicago Law Review* 66:373–436.
- Georgantzis, N., and D. Navarro-Martinez. (2010). Understanding the WTA-WTP Gap: Attitudes, Feelings, Uncertainty and Personality. *Journal of Economic Psychology* 31:895–907.
- Gilovich, T., and V. H. Medvec. (1995). The Experience of Regret: What, When, and Why. *Psychology Review* 102:379–95.
- Ginsburg, T., J. S. Masur, and R. H. McAdams. (2014). Libertarian Paternalism, Path Dependence, and Temporary Law. *University of Chicago Law Review* 81: 291–359.



- Hammack, J., and G. M. Brown Jr. 1974. *Waterfowl and Wetlands: Toward Bioeconomic Analysis*. Washington, DC: RFF Press.
- Harbaugh, W. T., K. Krause, and L. Vesterlund. (2001). Are Adults Better Behaved Than Children? Age, Experience, and the Endowment Effect. *Economics Letters* 70:175–81.
- Harless, D. W. (1989). More Laboratory Evidence on the Disparity between Willingness to Pay and Compensation Demanded. *Journal of Economic Behavior & Organization* 11:359–79.
- Hartman, R. S., M. J. Doane, and C. Woo. (1991). Consumer Rationality and the Status Quo. *Quarterly Journal of Economics* 106:141–62.
- Heffetz, O., and J. A. List. (2011). Is the Endowment Effect a Reference Effect? NBER Working Paper no. 16715.
- Horowitz, J. K., and K. E. McConnell. (2002). A Review of WTA/WTP Studies. *Journal of Environmental Economics and Management* 44:426–47.
- Hovenkamp, H. (1990). Marginal Utility and the Coase Theorem. *Cornell Law Review* 75:783–810.
- Hovenkamp, H. (1991). Legal Policy and the Endowment Effect. *Journal of Legal Studies* 20:225–47.
- Hovenkamp, H. (1994). The Limits of Preference-Based Legal Policy. *Northwestern University Law Review* 49:4–91.
- Isoni, A., G. Loomes, and R. Sugden. (2011). The Willingness to Pay–Willingness to Accept Gap, the “Endowment Effect,” Subject Misconceptions, and Experimental Procedures for Eliciting Valuations: Comment. *American Economic Review* 101:991–1011.
- Johnson, E., G. Haubl, and A. Keinan. (2007). Aspects of Endowment: A Query Theory of Value Construction. *Journal of Experimental Psychology* 33:461–74.
- Jones, O. (2001). Time-Shifted Rationality and the Law of Law’s Leverage: Behavioral Economics Meets Behavioral Biology. *Northwestern University Law Review* 95:1141–206.
- Jones, O., and S. F. Brosnan. (2008). Law, Biology, and Property: A New Theory of the Endowment Effect. *William and Mary Law Review* 49:1935–90.
- Kahneman D., J. Knetsch, and R. Thaler. 1990. Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy* 98:1325–48.
- Kahneman D., J. Knetsch, and R. Thaler. 1991. Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives* 5:193–206.
- Kelman, M. (1979). Consumption Theory, Production Theory, and Ideology in the Coase Theorem. *Southern California Law Review* 52:669–98.
- Kelman, M. (1998). Behavioral Economics as Part of a Rhetorical Duet: A Response to Jolls, Sunstein, and Thaler. *Stanford Law Review* 50:1577–92.
- Kennedy, Duncan (1981). Cost-Benefit Analysis of Entitlement Problems: A Critique. *Stanford Law Review* 33:387–445.
- Kingsley, D. C., and T. C. Brown. (2012). Does Prompting for Revision Influence Subjects’ Offers in Willingness to Accept–Willingness to Pay Lab Experiments? *Economics Bulletin* 32: 2580–85.
- Klass, G., and K. Zeiler. (2013). Against Endowment Theory: Experimental Economics and Legal Scholarship. *UCLA Law Review* 61:2–64.
- Knetsch, J. L. (1989). The Endowment Effect and Evidence of Nonreversible Indifference Curves. *American Economic Review* 79:1277–84.
- Knetsch, J. L., and J. A. Sinden. (1984). Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value. *Quarterly Journal of Economics* 99:507–21.

- Knetsch, J. L., and W. Wong. (2009). The Endowment Effect and the Reference State: Evidence and Manipulations. *Journal of Economic Behavior & Organization* 71:407–13.
- Korobkin, R. (1994). Note, Policymaking and the Offer/Asking Price Gap: Toward a Theory of Efficient Entitlement Allocation. *Stanford Law Review* 46:663–708.
- Korobkin, R. (1998a). The Status Quo Bias and Contract Default Rules. *Cornell Law Review* 83:608–87.
- Korobkin, R. (1998b). Inertia and Preference in Contract Negotiation: The Psychological Power of Default Rules and Form Terms. *Vanderbilt Law Review* 51:1583–651.
- Korobkin, R. (2000). Behavioral Analysis and Legal Form: Rules vs. Standards Revisited. *Oregon Law Review* 79:23–59.
- Korobkin, R. (2003). The Endowment Effect and Legal Analysis. *Northwestern University Law Review* 97:1227–91.
- Korobkin, R. (2009). Libertarian Welfarism. *California Law Review* 97:1651–85.
- Korobkin, R. (2011). What Comes after Victory for Behavioral Law and Economics? *University of Illinois Law Review* 2011:1653–74.
- Korobkin, R., and C. Guthrie. (1994). Psychological Barriers to Litigation Settlement: An Experimental Approach. *Michigan Law Review* 93:107–92.
- Kőszegi, B., and M. Rabin. (2006). A Model of Reference Dependent Preferences. *Quarterly Journal of Economics* 121:1133–65.
- Lakshminaryanan, V., M. K. Chen, and L. R. Santos. (2008). Endowment Effect in Capuchin Monkeys. *Philosophical Transactions: Biological Sciences* 363:3837–44.
- Lewinsohn-Zamir, D. (2012). Taking Outcomes Seriously. *Utah Law Review* 2012:861–902.
- List, J. A. (2003). Does Market Experience Eliminate Market Anomalies? *Quarterly Journal of Economics* 118:41–71.
- List, J. A. (2011). Does Market Experience Eliminate Market Anomalies? The Case of Exogenous Market Experience. *American Economic Review* 101:313–17.
- Loomes, G., and R. Sugden (1982). Regret Theory: An Alternative Theory of Rational Choice under Uncertainty. *Economics Journal* 92:805–24.
- Marshall, J. D., J. Knetsch, and J. A. Sinden. (1986). Agents' Evaluations and the Disparity in Measures of Economic Loss. *Journal of Economic Behavior and Organization* 7:115–27.
- McCaffery, E. J., D. J. Kahneman, and M. L. Spitzer. (1995). Framing the Jury: Cognitive Perspectives on Pain and Suffering Awards. *Virginia Law Review* 81:1341–420.
- Millon, D. (1998). Default Rules, Wealth Distribution, and Corporate Law Reform: Employment at Will versus Job Security. *University of Pennsylvania Law Review* 146:975–1041.
- Plott, C. R., and K. Zeiler. (2005). The Willingness to Pay–Willingness to Accept Gap, the “Endowment Effect,” Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. *American Economic Review* 95:530–45.
- Plott, C. R., and K. Zeiler. (2007). Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory. *American Economic Review* 97:1449–66.
- Polinsky, A. M. (1980). Resolving Nuisance Disputes: The Simple Economics of Injunctive and Damage Remedies. *Stanford Law Review* 32:1075–112.
- Posner, R. A. (2007). *Economic Analysis of Law*. 7th ed. Austin, TX: Wolters Kluwer for Aspen Publishers.
- Public Law 112–240. The American Taxpayer Relief Act of 2012. 126 Stat. 2313.
- Rachlinski, J. J. (1996). Gains, Losses, and the Psychology of Litigation. *Southern California Law Review* 70:113–85.

- Rachlinski, J. J. (2000). The Psychology of Global Climate Change. *University of Illinois Law Review* 2000:299–320.
- Rachlinski, J. J., and F. Jourden. (1998). Remedies and the Psychology of Ownership. *Vanderbilt Law Review* 51:1541–82.
- Ritov, I. and J. Baron. (1990). Reluctance to Vaccinate: Omission Bias and Ambiguity. *Journal of Behavioral Decision Making* 3:263–77.
- Rowe, R. D., R. C. D'Arge, and D. S. Brookshire. (1980). An Experiment on the Economic Value of Visibility. *Journal of Environmental Economics and Management* 7:1–19.
- Rutherford, M., J. L. Knetsch, and T. C. Brown. 1998. Assessing Environmental Losses: Judgments of Importance and Damage Schedules. *Harvard Environmental Law Review* 22:51–101.
- Samuelson, W., and R. Zeckhauser. (1988). Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty* 1:7–59.
- Sayman, S., and A. Onculer. (2005). Effects of Study Design Characteristics on the WTA-WTP Disparity: A Meta Analytical Framework. *Journal of Economic Psychology* 26:289–312.
- Shogren, J. F., S. Y. Shin, D. J. Hayes, and J. B. Kliebenstein. (1994). Resolving Differences in Willingness to Pay and Willingness to Accept. *American Economic Review* 84:255–70.
- Simonson, I., and A. Drolet. (2004). Anchoring Effects on Consumers' Willingness-to-Pay and Willingness-to-Accept. *Journal of Consumer Research* 31:681–90.
- Stake, J. E. (2004). The Property "Instinct." *Philosophical Transactions: Biological Sciences* 359:1763–74.
- Strahilevitz, M. A., and G. Loewenstein. (1998). The Effect of Ownership History on the Valuation of Objects. *Journal of Consumer Research* 25:276–89.
- Sunstein, C. R. (1986). Legal Interference with Private Preferences. *University of Chicago Law Review* 53:1129–74.
- Sunstein, C. R. (1998). Selective Fatalism. *Journal of Legal Studies* 27:799–823.
- Taylor, S. E. (1991). Asymmetrical Effects of Positive and Negative Events: The Mobilization-Minimization Hypothesis. *Psychological Bulletin* 110:67–85.
- Thaler, R. (1980). Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior and Organization* 1:39–60.
- Tversky, A., and D. Kahneman. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics* 106:1039–61.
- United States v. Carroll Towing Co.*, 159 F.2d 169 (2d Cir. 1947).
- van Dijk, E., and D. van Knippenberg. (1998). Trading Wine: On the Endowment Effect, Loss Aversion, and the Comparability of Consumer Goods. *Journal of Economic Psychology* 19:485–95.
- Weaver, R. (2008). Three Essays in Decision Making: Transaction Disutility and the Endowment Effect. Ph.D. diss., Massachusetts Institute of Technology.
- Weaver, R., and S. Frederick. (2012). A Reference Price Theory of the Endowment Effect. *Journal of Marketing Research* 49:696–707.
- Williams, D. R. (1995). Valuing Natural Environments: Compensation, Market Norms, and the Idea of Public Goods. *Connecticut Law Review* 27:365–491.
- Willis, Lauren. (2013). When Nudges Fail: Slippery Defaults. *University of Chicago Law Review* 80:1155–1230.
- Zamir, E. (1997). The Inverted Hierarchy of Contract Interpretation and Supplementation. *Columbia Law Review* 97:1710–803.

Zamir, E. (2012). Loss Aversion and the Law. *Vanderbilt Law Review* 65:829–94.

Zamir, E., and I. Ritov. (2010). Revisiting the Debate over Attorneys' Contingent Fees: A Behavioral Analysis. *Journal of Legal Studies* 39:245–88.

Zamir, E., and I. Ritov. (2012). Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation. *Journal of Legal Studies* 41:165–207.

## CHAPTER 13

---

# PROBABILITY ERRORS

## *Overoptimism, Ambiguity Aversion, and the Certainty Effect*

---

SEAN HANNON WILLIAMS

### 1 INTRODUCTION

---

A huge body of literature now documents numerous deviations from expected utility theory. For example, the availability heuristic and the affect heuristic can cause people to overestimate risks that are easy to recall or especially salient. Other chapters have dealt with these biases. This chapter discusses three other species of probability errors: overoptimism, ambiguity aversion, and the certainty effect. It provides a brief introduction to each of these phenomena and offers a representative sampling of how they can be used to inform our understanding of human behavior and assist policymakers in designing legal reforms.

SECTION 2 of this chapter introduces the vast literature on overoptimism and its impact on legal scholarship. Overoptimism causes people to underestimate their likelihood of negative events, to overestimate the likelihood of positive events, and to be overly confident in each of these erroneous judgments. It also distorts people's reactions to explicit and accurate probabilistic information that they might obtain from others; informing people of their actual risk of heart attack, for example, is unlikely to have as much impact on their prior probability estimates as it should. Despite the existence of both situational and dispositional factors that can mitigate overoptimism, a huge range of people exhibit overoptimism in a huge range of situations. For example, overoptimism affects students when they assess their level of knowledge, leadership skills, athleticism, generosity, and civic responsibility. It affects adults in the general population when they assess their driving skill, health risk, ability to pick stocks, and the quality of their marriages. It even affects practicing lawyers when they try to predict the outcome of their cases. Following an exposition of the psychological findings in section 2.1, section 2.2 provides an overview of the myriad ways that legal scholars have used

overoptimism to better understand human behavior. Overoptimism helps explain a host of behavioral patterns, including consumer borrowing behavior, the prevalence of poor decision-making around litigation and settlement in both the civil and criminal contexts, the content of liquidated damages clauses in contracts, the investment decisions of laypersons and CEOs, and the failure of many of us to adequately prepare for a host of negative events.

SECTION 3 discusses the second and third phenomena that this chapter is centered around: ambiguity aversion and the certainty effect. Both ambiguity aversion and the certainty effect affect how people respond to explicit and accurate probabilistic information. Ambiguity aversion reflects the tendency of people to prefer known to unknown risks. That is, people prefer to take a gamble with a 15% chance of losing over a gamble with a 10 to 20% chance of losing. Section 3.1 outlines the empirical evidence supporting ambiguity aversion. Ambiguity aversion has been found in studies of students, members of the general population, actuaries, insurance underwriters, deliberating partners, and even monkeys. Legal scholars have used ambiguity aversion to help explain the stickiness of default rules in contract law, the use of so-called “material adverse effect” clauses, and various aspects of plea-bargaining behavior. Legal scholars have also explored the possibility of strategically inserting or highlighting ambiguity within the law in order to influence behavior and, for example, increase compliance with the tax code. Section 3.2 provides an overview of the certainty effect. People who exhibit a certainty effect place a premium on certainty and are willing to pay a great deal to reduce the probability of negative events to zero. The certainty effect helps explain numerous phenomena, including jurors’ resistance to circumstantial evidence, the undesirability of probabilistic insurance, and various patterns in both environmental statutes and agency enforcement of those statutes.

ALL three of the phenomena discussed here have dramatically enhanced our understanding of human behavior and will continue to play an important role in predicting the effects of legal and policy choices.

## 2 OVEROPTIMISM

---

### 2.1 Empirical Findings

Overoptimism—an umbrella term that covers a class of more specific biases—is common, costly, and difficult to correct. People routinely believe that they are better than most of their peers. This is commonly referred to as the better-than-average effect or the above-average effect. People also exhibit overconfidence, and thereby have an inaccurate degree of confidence in their abilities as measured on objective scales. People also fall prey to a host of self-serving biases that cause them to interpret ambiguous information in ways that cast them in a favorable light. These forms of overoptimism create two interrelated problems. First, overoptimism can cause people to generate

erroneous probability estimates. For example, people routinely underestimate their risk of heart attacks and various other negative health events. Second, overoptimism affects how people respond to debiasing information that provides them with explicit and accurate probability estimates. For example, if the government provides information about average heart attack risks, people are likely to discount this information because of their belief that they are less likely to suffer a heart attack than the average person.

The simplest measure of overoptimism reveals a strong above-average effect. The easiest way to test a population for overoptimism is to ask people to compare themselves to their peers. If people were accurate in their self-assessments, then about half would rate themselves as above average (assuming a symmetrical distribution of the relevant trait around a population mean). However, research routinely finds that the vast majority of subjects believe that they are above average (Alicke and Govorun 2005; Weinstein 1980). This is true both for college students—the subjects of much psychological research—and for members of the general population (Jolls, Sunstein, and Thaler 1998; Weinstein 1987). Compared to their peers, people overwhelmingly think that they are better drivers, better leaders, better managers, and better workers. They also believe they are smarter, healthier, more sensitive, more ethical, more charitable, more athletic, more attractive, more likely to vote, and more productive. These above-average effects are endemic in America, Canada, and Western Europe (Dunning, Heath, and Suls 2004). For example, 94% of American college professors think that their work is above average, and 25% of American students think that they are in the top 1% of their peers in terms of leadership ability (Williams 2008).

People exhibit a similar tendency when asked to assess their comparative risk of suffering harm. People think that they are less likely than their peers to experience a host of negative life events including alcoholism, car accidents, divorces, unemployment, unwanted pregnancy, criminal victimization, heart attacks, heart disease, strokes, and skin cancer (Williams 2008). Put another way, people believe that they have above-average skill at avoiding most negative events.

Although the simplest way to measure overoptimism is to ask people to compare themselves to their peers, researchers have also measured overoptimism by comparing self-assessments with objective measurements of ability. When subjects report being 90% sure that their answer to a general knowledge question on a quiz was correct, we might suspect that such answers will be correct about 90% of the time. But these answers are only correct between 40% and 75% of the time (Bolger and Önköl-Atay 2004). Similar patterns emerge when people predict how long it will take them to complete a task. This planning fallacy is both common and substantial. Using other forms of objective assessments, research has shown that people exhibit overly rosy beliefs about their athletic prowess, job performance, and even their ability to detect lies. Overall, self-assessments only have a mild correlation with actual performance (Dunning, Heath, and Suls 2004).

Overoptimism is particularly persistent over time because of two motivational biases—the self-serving bias and the attribution bias—that lead people to



systematically misinterpret data when they have some emotional stake in the outcome (Chambers and Windschitl 2004; Weinstein and Klein 1995; Dunning, Meyerowitz, and Holzberg 1989). These biases affect a host of decisions, only some of which relate to overoptimism. For example, people often make self-serving assessments about the fairness of a particular situation (Loewenstein and Moore 2004). But these biases also help create and maintain overly optimistic probability estimates.

The self-serving bias leads people to interpret ambiguous information in ways that support their desired outcome. “Instead of a naive scientist entering the environment in search of the truth, we find the rather unflattering picture of a charlatan trying to make the data come out in a manner most advantageous to his or her already-held theories” (Fiske and Taylor 1984). In one set of studies, subjects were assigned the role of plaintiff or defendant in a tort suit and given information about the case. Subjects who were assigned the role of the plaintiff placed more weight on the facts that yielded a higher award. Similarly, subjects assigned the role of defendant assigned more importance to the facts that supported a lower award. Plaintiffs also predicted higher awards than defendants (Babcock and Loewenstein 1977). These optimistic predictions are not limited to student subjects. Self-serving biases have also been found among lawyers and professional negotiators (Babcock and Loewenstein 1977). Ironically, increasing the amount of information available to the decision-maker can strengthen rather than weaken the self-serving bias (Thompson and Loewenstein 1992; Dunning, Heath, and Suls 2004). Increasing the amount of information will often increase the number of potentially plausible interpretations of that information. A consumer who knows that she has borrowed a lot in the past may not predict that she will borrow in the future because she has enough information about her past situation to convince herself that her past borrowing was abnormal and unlikely to repeat itself. Just as a person who does poorly on a test will tend to view it as uninformative (Dogan et al. 2011), she will discount the relevance of her past borrowing. After all, she is a responsible person who would never live beyond her means.

The attribution bias leads people to attribute negative outcomes to bad luck, while attributing positive outcomes to their own personal acumen (Mezulis et al. 2004). Consider again a consumer who has a great deal of private information about the circumstances of her past borrowing. This consumer will attribute her past borrowing to bad luck. It was just bad luck that several bills arrived at once, and it certainly was not the result of poor planning or poor self-control. Because people can interpret their past borrowing behavior as a response to unique circumstances rather than durable personality traits, they will do so. This hinders learning from experience.

One student study painted a particularly stark picture of the way that people distort information when making predictions of their future performance. When students predicted the performance of their peers on an exam, they relied heavily (and correctly) on past exam scores. But when they made predictions for themselves, they discounted past performance information and gave undue weight to their aspirations—the scores that they hoped to achieve. A similar asymmetry emerged when students decided what type of information was most important for making these predictions. When students

were asked what information they would give to someone else to help them predict the student's future test score, they consistently preferred to supply them with their hoped-for score rather than their past scores. In contrast, when students were requesting information from others, they consistently ranked past exam score information as more important than the other person's mere aspirations (Helzer and Dunning 2012). Here, as in other areas, people show "the triumph of hope over experience" (Boswell 1874).

Self-serving biases, the attribution bias, and other forms of motivated reasoning make overoptimism resilient to debiasing (Weinstein 1995; Weinstein 1980). These barriers to debiasing persist even when people make repeated decisions with repeated feedback, and even if they are given incentives to make accurate predictions. People selectively misremember negative feedback and process negative information shallowly (Sedikides and Green 2004; Sanitioso and Wlodarski 2004). In one study of auto accidents, researchers compared the self-assessed driving skill of 50 people without any accident history with 50 people who had recently been hospitalized because of a car accident—34 of whom had been at fault. Perhaps unsurprisingly, these two groups both felt that their driving skill was above average. More surprisingly, the two groups did not differ with respect to their self-assessed driving skill (Preston and Harris 1965). One would have hoped that at least some of the drivers who were at fault would have incorporated this evidence into their skill assessments. A study of bridge players at an Illinois bridge club is also informative. Although the players accurately recalled their record within the club, they were overconfident when predicting the result of their next game despite being offered rewards (albeit small ones) for accuracy. This overconfidence persisted, unchanged in magnitude, across more than five predictions each of which was followed by outcome feedback (Simons 2013). The same pattern appears in student populations. In one study, students in an economics class took three tests and predicted their performance on each. After each test they received feedback on their performance (in addition, of course, to the performance feedback they already received in past economics classes, and on past exams). All students were overly confident on the first exam. Only the students who did best on the exams showed any signs of reducing their overconfidence over time. Student who got below-average scores on the exams did not become more accurate, despite monetary incentives for accuracy (Ferraro 2010).

Even successful debiasing can be short lived and somewhat domain specific. For example, immediately after the 1989 California earthquake, students were not overly optimistic about their likelihood of being injured in a natural disaster. They were, however, still overly optimistic in their predictions about other risks. After three months, those students returned to being overly optimistic even about natural disasters (Dunning, Heath, and Suls 2004).

Although overoptimism is widespread and resilient, it is not omnipresent. Both dispositional and situational factors affect its pervasiveness and severity. As a person's perceived control increases, so too does her overoptimism (Dunning, Heath, and Suls 2004). If people are assured that a potential negative event is completely outside of their control, for example if its occurrence depends on another person's roll of a set of dice,

people exhibit minimal overoptimism (Camerer and Lovallo 1999). But when people have even a very small amount of control over the negative event, such as when they roll the dice themselves, they will exhibit at least some unrealistic optimism (Sedikides and Gregg 2008). The threat of immediate, accurate, and unambiguous feedback can also reduce overoptimism (Armor and Sackett 2006). For any given set of situational constraints, a person's level of optimism is also affected by dispositional factors. Young men, for example, tend to be especially optimistic across a host of domains (Barber and Odean 2001). And of course, some people are just more optimistic than others (Simons 2013).

Similar patterns are nicely illustrated by the above-average effect, which is largest for risks that are relatively controllable and rare, and in persons who are high in self-esteem. For some risks, subjects can even exhibit a below-average effect (Chambers and Windschitl 2004). For example, populations might show above-average effects when asked whether they will live beyond age 70, but below-average effects when asked whether they will live beyond age 100. Below-average effects can also occur for especially hard tasks. In one study, students believe that they will beat their peers in a trivia contest when the topic is TV sitcoms, but believe that they will lose when the topic is Mesopotamia (Moore 2007). Despite these occasional below-average effects, above-average effects are much more prevalent (Sedikides and Alicke 2012). Across many common and important decisions like whether to buy insurance or whether to exercise more to reduce one's risk of a heart attack, the above-average effect is likely to be dominant.

Overoptimism is not always problematic; it has numerous benefits. It helps maintain happiness, self-esteem, and relationship satisfaction (Mezulis et al. 2004; Srivastava et al. 2006). Optimistic people are also physically healthier; they have speedier recoveries, fewer symptoms, and cope better with mild health problems (Scheier and Carver 1985). Overoptimism can also correct other biases. Overoptimism often leads people to underestimate risks. Other biases—such as the availability heuristic, the affect heuristic, and affective forecasting errors—can cause people to overestimate risks. If a particular situation triggers offsetting biases, then people might come to a roughly accurate risk assessment (Jolls and Sunstein 2006). Nonetheless, we often have data that suggests that, in particular circumstances and for particular populations, one of the biases is swamping the other and people are misestimating the relevant risks (Williams 2008).

## 2.2 Applications

The remainder of this section contextualizes overoptimism by analyzing its effects in specific decision-making circumstances. Overoptimism touches virtually all aspects of decision-making. This is evident in the number of other chapters in this *Handbook* that use overoptimism to illuminate substantive controversies. Oren Bar-Gill's chapter argues that overoptimism plays an important role in consumer borrowing (chapter 18 by Bar-Gill in this volume). Consumers are likely to overestimate the likelihood that

they will be able to pay off a debt, while simultaneously underestimating the likelihood that problems such as heart attacks, divorce, and car accidents will force them to take on even more debt (Bar-Gill 2004). These dynamics affect many avenues of consumer borrowing, such as credit cards, payday loans, and subprime mortgages (Bar-Gill and Warren 2008; Francis 2010). Jennifer Robbennolt's chapter illustrates how a host of phenomena—including overconfidence and self-serving biases—impede settlement by leading lawyers and their clients to misjudge the expected value of their case. She argues that lawyers could likely benefit from implementing various debiasing methods. For example “consider the opposite” exercises have shown some, albeit not uniform, success at reducing the effects of overoptimism (chapter 24 by Robbennolt in this volume; Goodman-Delahunty et al. 2010). Russell Covey's chapter focuses on defendants in criminal trials, and observes that overoptimism leads these people to overestimate their likelihood of prevailing at trial. Self-serving biases exacerbate this basic tendency because both prosecutors and defendants will often reach differing conclusions based on the same objective data. Covey argues that several seemingly oppressive aspects of the criminal justice system might help mitigate the effects of this overoptimism in the context of plea bargaining (chapter 25 by Covey in this volume). Melvin Eisenberg's chapter notes that overoptimism leads contracting parties to pay insufficient attention to liquidated-damages provisions (chapter 17 by Eisenberg in this volume). Kent Greenfield's chapter discusses the effects of overoptimism on CEOs, investors, and analysts (chapter 20 by Greenfield in this volume). Of course, overoptimism is evident in many other areas as well. This section briefly outlines two additional areas where overoptimism is likely to be prevalent, costly, and difficult to correct: healthcare and marriage.

### 2.2.1 *Healthcare*

Overly optimistic people are likely to take inadequate precautions regarding a host of health-related risks. If people made more realistic estimates of their health risks, it is likely that they would do more to avert negative health events. Drawing on popular images from the Obamacare debate, one might say that people who accurately assessed their health risks would not need the government to force them to eat broccoli; they would eat more broccoli themselves.

Overly optimistic perceptions of health risks can be hard to correct. Mandatory disclosures that only reveal population averages are unlikely to be effective. Any time people are presented with risk information in the form of population averages, the above-average effect undermines the usefulness of the message. Revealing that the average adult has a certain risk of stroke may lead people to think that strokes are a large problem overall, but will not necessarily lead people to feel vulnerable themselves.

Informing people about risk factors also has limits. In one illustrative study, middle-aged people who believed they had a below-average risk of having a heart attack exhibited two forms of resistance to debiasing information. First, when given the choice to read about one of several risk factors, overly optimistic subjects chose to read more about a risk factor that offered them good news. So a smoker with low

cholesterol might read more about the link between cholesterol and heart attacks, and pay less attention to the link between smoking and heart attacks. Second, when overly optimistic subjects were asked to read about a host of risk factors, they retained less of the information than their more realistic peers (Radcliffe and Klein 2002; Rothman and Kiviniemi 1999). More subtle interventions have had more success (Chandler et al. 1999). Self-serving effects can be reduced if the patient's self-worth is validated before receiving the information. For example, reminding people of a time when they were especially kind can make them less likely to resist health risk information (Dunning, Heath, and Suls 2004).

Even personalized disclosures can fail to alter risk assessments that are skewed by overoptimism. In one early study, people from the general population were asked to rate their relative risk of having a heart attack, having a stroke, getting into a fatal car accident, and getting cancer. In order to create a more accurate measure of their relative risk, they also answered questions about their health, behaviors, family history of disease, and so on. Using this data, researchers constructed actuarial estimates of their risk. Not surprisingly, people exhibited overoptimism. More surprisingly, people did not consistently correct their overly optimistic beliefs even after being informed about their individualized risk estimate. For some risks, optimism decreased. But for others, risk estimates were completely unaffected, and some people actually became more optimistic when presented with information that should have been a wake-up call about their comparative health risk (Kreuter and Strecher 1995; Gerrard, Gibbons, and Reis-Bergan 1999).

Encouraging people to procure health insurance is one potential response to these persistent errors. Insurance reduces the costs of healthcare should one's health deteriorate, and reduces the probability of this deterioration by increasing the likelihood of treating health problems early. But the fact that people underestimate their risk of a host of major medical calamities (Weinstein 1987) suggests that many people undervalue health insurance. In response to this possibility, Tom Baker has suggested fighting fire with fire—or here, overoptimism with overoptimism—by bundling health insurance with lotteries. Overly optimistic people will undervalue the insurance because they underestimate their health risks, but will overvalue the insurance because they will overestimate their likelihood of winning the prize (Baker and Siegelman 2010). Although precisely calibrating these two offsetting effects of overoptimism presents challenges, doing so offers the hope of at least reducing the effects of the pervasive tendency of people to underestimate their risk of suffering from common negative health events like heart attacks.

### 2.2.2 *Marriage*

In addition to exhibiting overoptimism about health outcomes, people are optimistic about marriage and especially unlikely to prepare for the possibility of divorce. Most people are overly optimistic about the probability that they will live happily ever after with their partner. Although people believe that 50% of marriages end in divorce, most also predict that their chance of divorce is 0%. Even those who acknowledge the

possibility of divorcing believe that their probability of divorce is drastically lower than average. For example, members of the general population estimated their personal likelihood of divorce at 10%. Law students—who are trained to see the worst-case scenario in legal relationships—are only slightly more realistic. They estimate their chances of divorce to be 17% (Mahar 2003; Baker and Emery 1993).

In addition to being overly optimistic about the probability of divorce, people are also overly optimistic about the financial effects of divorce. Although people predict that courts award alimony in 40% to 50% of cases, 80% of them predict that a court will award alimony in their case. Although people predict that 20% of obligees will never be able to collect a penny of their court-ordered support payments, 100% of them believe that their ex-spouse would dutifully pay every cent (Baker and Emery 1993). This pattern of overoptimism strongly suggests that people underprepare for the possibility of divorce, and might, for example, unwisely let their job skills atrophy in reliance on future alimony.

Marital overoptimism creates huge financial burdens on many families. These burdens fall disproportionately on women and children. When one spouse takes time off from her career to raise children, her long-term earning capacity often suffers. People are more likely to make these risky choices if they erroneously think that they will never divorce or that divorce law will provide them with an adequate safety net in the form of alimony and child support. When couples actually divorce, women who have made choices that reduce their earning capacity are left with a drastically lower standard of living and are often impoverished. Their husbands, in contrast, often have a higher standard of living after the divorce (Daniels et al. 2006).

This pattern of costly marital overoptimism could justify legal interventions. For example, prenuptial and postnuptial agreements might merit greater procedural or even substantive protections than general contracts because fiancés and spouses might underestimate the possibility of divorce. Divorce law might also include a robust alimony regime. This would insulate overly optimistic spouses from the full costs of their biases, while arguably doing a better job of reflecting the bargain that the parties thought they struck when they married. After all, most people appear to believe that a court would award alimony if they were to divorce (Williams 2008).

### 3 AMBIGUITY AVERSION AND THE CERTAINTY EFFECT

---

As discussed in the previous section, overoptimism can affect how people respond to explicit probability information. Ambiguity aversion and the certainty effect also affect how people respond to such information. Ambiguity aversion refers to the tendency of people to prefer known risks to unknown risks. That is, people would be more comfortable taking a risk described as having a 15% chance of a negative outcome than they



would taking a risk described as having a 10% to 20% chance of that outcome. The certainty effect describes the tendency of people to focus more on the *possibility* of a salient outcome rather than its (potentially low) *probability*. Accordingly, people are willing to pay a great deal to completely eliminate even minor risks. It is not immediately obvious that ambiguity aversion and the certainty effect reflect errors. A preference to avoid ambiguity could be akin to rational risk aversion, and certainty may have psychological benefits that justify giving it special weight. Nonetheless, there are strong reasons to think that each phenomenon is at least partially the result of biases.

### 3.1 Ambiguity Aversion

Ample evidence suggests that people prefer known risks to unknown risks. In a classic study, Daniel Ellsberg presented subjects with two urns. One contained 50 red balls and 50 black balls. The other contained red and black balls in unknown proportions. Subjects preferred to bet on the 50-50 urn, and avoided the urn with unknown numbers of red and black balls (Ellsberg 1961). This phenomenon has been termed ambiguity aversion.

People exhibit ambiguity aversion even when the ambiguity is small and the ambiguous option has a higher expected value. In Ellsberg's two-urn experiment, subjects faced a great deal of ambiguity. The second urn contained somewhere between 0 and 100 red balls. But subjects exhibit ambiguity aversion even when the relevant ambiguity is reduced. For example, people would prefer gambles with a 20% chance of winning to gambles with a 10% to 30% chance of winning even if they were told that all of the probabilities within that range are equally likely. Subjects avoid such ambiguous options even when the odds favor the ambiguous option, or when the ambiguous option offers a higher payoff (Keren and Gershten 1999; Trautmann, Vieider, and Wakker 2008). A significant number of subjects exhibit ambiguity aversion, and they pay substantial premiums to avoid ambiguity (Camerer and Weber 1992).

Ambiguity aversion is not merely a phenomenon that occurs in stylized lab experiments with student subjects. It has been detected among actuaries, insurance underwriters, deliberating partners, and even monkeys (Cabantous 2007; Keller, Sarin, and Souderpandian 2007; Kunreuther et al. 1995; Hayden et al. 2010). It appears in experimental markets, and data from real markets is consistent with ambiguity aversion (Aloysius 2005; Bossaerts et al. 2010).

Of course, ambiguity aversion is not universal. Studies have found a good deal of heterogeneity in subjects' response to ambiguity, but some patterns emerge.

Ambiguity aversion is stronger and more consistently found in the realm of gains than the realm of losses. In the realm of gains, people are consistently averse to ambiguity. In the realm of losses, a more complex pattern emerges. For high-probability losses, people are slightly ambiguity seeking rather than ambiguity averse. For example, when North Carolina businesspersons were asked whether they would choose to face an unambiguous 70% risk of loss or a 50%–90% risk of the same loss, they preferred



the ambiguous option. For low- and mid-probability losses, however, people remain averse to ambiguity (Hogarth and Einhorn 1990; Viscusi and Chesson 1999; Ho et al. 2002).

Ambiguity aversion is affected by the subject's perceived competence, which can be influenced by both situational and dispositional factors. When a particular situation makes people feel especially ignorant, they exhibit more ambiguity aversion (Fox and Tversky 1995). This occurs, for example, when people are primed to think about more knowledgeable people, or when people are overloaded with information that they do not know how to use (Du and Budescu 2005; Fox and Weber 2002). In contrast, when decision-makers feel that they are familiar with the relevant decision domain, they are less averse to ambiguity and can even become ambiguity seeking (Fox and Weber 2002).

Ambiguity aversion is also affected by the subject's level of optimism. Highly optimistic subjects are less likely to be averse to ambiguity. Subjects with low levels of optimism are more likely to exhibit ambiguity aversion (Pulford 2009).

Other aspects of the decision-maker and the decision environment can also affect ambiguity aversion. People are less averse to ambiguity when they know they are making repeated gambles, when the relevant ambiguity is less salient, and when they are higher in numeracy (Liu and Colman 2009; Fox and Tversky 1995; Halevy 2007). People are more averse to ambiguity when they anticipate that they may have to defend their judgment to others (Trautmann, Vieider, and Wakker 2008).

It is possible to characterize ambiguity aversion as a rational preference akin to risk aversion. However, the empirical evidence suggests that ambiguity aversion is at least partially the result of an error. More specifically, ambiguity aversion is related to the inability or unwillingness to reduce compound lotteries (Halevy 2007).

Regardless of whether ambiguity aversion is best characterized as a bias or a preference, it can help explain behavioral patterns that are relevant to several areas of law. Omri Ben-Shahar and John Pottow have argued that ambiguity aversion helps explain the stickiness of default terms in contract law, even when those default terms could be altered in mutually beneficial ways. When one party suggests a deviation from a default term, the other party is exposed to greater ambiguity, and is also perhaps more likely to wonder whether they are comparatively ignorant of the new term's implications. Therefore, ambiguity aversion helps explain why people are "spooked" by nonstandard contract terms (Ben-Shahar and Pottow 2006). Amitai Aviram and Avishalom Tor have similarly argued that ambiguity aversion will decrease the likelihood that companies would enter into novel information-sharing agreements, and push them toward more familiar competitive practices (Aviram and Tor 2003). Eric Talley has argued that extreme ambiguity about the potential circumstances that could affect the value of a contract leads people away from merely adjusting the contract price and toward including "material adverse change" clauses that allow one party to escape the contract entirely if the uncertain event occurs (Talley 2009). In the realm of criminal law, Uzi Segal and Alex Stein have identified what they call asymmetric ambiguity aversion in plea bargaining. The defendant is averse to the ambiguity involved in a jury trial, while

the prosecution, being a repeat player, is not as averse to ambiguity. They suggest ways to reduce ambiguity in order to reduce the resulting disparity in bargaining power. For example, if the outcome of bench trials is more predictable and hence less ambiguous than the outcome of jury trials, and defendants could unilaterally opt for a bench trial, then the asymmetry that they identify could be reduced (Segal and Stein 2006).

Ambiguity aversion can also be used strategically. Consider a questionable tax deduction. If it is harder to predict when you might be subject to a tax penalty for erroneously using the deduction—that is, if the likelihood of incurring a penalty is more ambiguous—then ambiguity-averse people will be less likely to take the deduction. In general, increased ambiguity could increase the deterrent effect of legal penalties. In this and other ways, it may be possible to increase tax compliance by highlighting ambiguity in the tax code (Lawsky 2013). Ambiguity can also be used strategically to reduce distortions caused by the tax code. Yuval Feldman and Shahar Lifshitz have argued that ambiguous tax benefits will have a smaller *ex ante* effect on behavior than nonambiguous ones. While people may try to alter their behavior to fit into a tax loophole if they are sure that it will provide benefits, they will be significantly less inclined to do so if the probability of the benefit is ambiguous (Feldman and Lifshitz 2011).

### 3.2 The Certainty Effect

When people respond to known probabilities, they are often much more responsive to the possibility of an event than its probability. When people are asked about their willingness to pay to reduce a risk, they are willing to pay far more when the risk is eliminated. Suppose that someone is willing to pay \$100 to reduce her risk of dying from 3 in 100,000 to 2 in 100,000. She would be willing to pay far more to reduce her risk from 1 in 100,000 to 0, even though in each scenario she is asked for her willingness to pay for a risk reduction of 1 in 100,000. Research into this certainty effect finds that people are more sensitive to changes in probability that result in eliminating the possibility of an outcome or making it certain to occur. Conversely, people tend to be relatively insensitive to changes in probability that do not result in making the relevant outcome either certain or impossible.

Early evidence for the certainty effect came from Maurice Allais and what became known as the Allais paradox. Allais presented subjects with the choice between two high-reward gambles similar to the following two. The first offered an 11% chance of winning \$1,000,000 with a complementary 89% chance of winning nothing. The second offered a 10% chance of winning \$5,000,000 and a complementary 90% chance of winning nothing. Subjects chose the second bet. But when each gamble was altered by adding an 89% chance of winning \$1,000,000, subjects reversed their preferences and chose the first gamble. Why? The addition of an 89% chance of winning \$1,000,000 made the first “gamble” a seductive 100% chance of receiving \$1,000,000. Subjects preferred certainty to taking the alternate gamble of an 89% chance of \$1,000,000, a 10% chance of winning \$5,000,000, and a 1% chance of winning nothing (Allais 1953). This

paradox has been replicated with much smaller gambles, although the certainty effect appears to be stronger when larger amounts are at stake. Allais's paradox has even been found when people choose between different commuting routes that offer various probabilities of various commute times (Avineri and Prashker 2004).

The certainty effect also helps explain why subjects would choose to receive \$300 with certainty over an 80% chance of receiving \$400. This preference is not merely a preference for the option with the higher probability; it is a preference for certainty. When the probabilities are multiplied by a common factor of  $1/10$ —yielding a choice between a 10% chance of receiving \$300 and an 8% chance of receiving \$400—subjects reverse their initial preference and choose the lower chance of receiving more money. This “common ratio effect” is a form of the certainty effect and lends further credence to the claim that certainty plays a special role in decision-making (Baucells and Heukamp 2010).

The above studies suggest that people overvalue certainty. Other studies complement this finding by suggesting that people undervalue probability changes that do not result in certainty. For example, Amos Tversky and Daniel Kahneman found that the median subject was willing to pay \$10 for a lottery ticket that provided a 1% chance of winning \$200. They were willing to pay \$188 for a lottery ticket that provided a 99% chance of winning \$200. Thus, subjects valued the first percentage point (the move from impossibility to possibility) at \$10, and valued the last percentage point (the move from possibility to certainty) at \$12. This implies that subjects valued the 98 intermediate percentage points at \$178, or a mere \$1.80 each. A similar pattern emerged when subjects indicated their willingness to pay to avoid potential losses (Tversky and Kahneman 1992).

The certainty effect is more pronounced when the relevant outcome is more emotionally salient. In one study, subjects were asked about their willingness to pay to avoid a 1% and 99% chance of losing \$20. Subjects were also asked their willingness to pay to avoid a 1%, 99%, and 100% chance of a “short, painful, but not dangerous electric shock.” When subjects encountered the monetary loss condition, their responses were sensitive to the relevant probabilities. They were willing to pay about \$1 to eliminate the 1% risk, and \$18 to eliminate the 99% risk. In contrast, subjects were far less sensitive to probability in the electric shock condition. Subjects were willing to pay \$7 to avoid a 1% chance of being shocked, \$10 to avoid a 99% chance, and about \$20 to avoid a certain shock (Rottenstreich and Hsee 2001). Two main lessons emerge from this study. First, people are willing to pay a large amount to avoid even a small probability of an emotionally salient risk. Second, people are relatively insensitive to the probability of salient risks. Here, subjects were relatively insensitive to the difference between a 1% chance and a 99% chance of getting shocked; they implicitly valued these 98 intermediate percentage points at \$3, or a mere 3.06 cents each. Again, people focus disproportionately on the mere possibility of emotionally salient risks, rather than their probability (Sunstein 2003).

The certainty effect is not ubiquitous. Some research suggests that it is reduced or eliminated when payouts are delayed. One study presented subjects with a choice

between receiving \$9 with certainty and a gamble with an 80% chance of yielding \$12. When payouts were immediate, subjects exhibited a certainty effect. But subjects preferred the gamble when the payouts were delayed for three months. Further increasing the delay further decreased the certainty effect (Baucells and Heukamp 2010). Other studies have found more mixed results (Weber and Chapman 2005). Nonetheless, this result is consistent with a large body of psychological literature arguing that people process information about temporal distance, spatial distance, social distance, and risk in similar ways (Liberman and Trope 2008). Therefore, adding a delay is treated as if it added risk. The presence and magnitude of the certainty effect is also influenced by personal characteristics of the decision-maker. For example, in one Dutch study more educated subjects were less likely to exhibit a certainty effect and more likely to make decisions consistent with expected utility theory (Huck and Muller 2012).

The certainty effect can help explain a number of puzzles. Kevin Jon Heller has argued that the certainty effect helps explain jurors' resistance to circumstantial evidence. The prospect of falsely convicting a defendant is particularly salient. Therefore, jurors exhibit a strong certainty effect and discount probabilistic evidence even if it purports to show a 99% chance of guilt. Instead, jurors want evidence that has an appearance of certainty, such as a confession or the testimony of an eyewitness who swears that the defendant committed the crime (Heller 2006).

The certainty effect also helps explain the undesirability of probabilistic insurance. Imagine an insurance contract that provided that, in the event that you suffered a loss, the insurance company would flip a coin and only reimburse you if it landed on heads. Subjects demand huge reductions in premiums in order to face this form of probabilistic insurance. In one study by Peter Wakker, Richard Thaler, and Amos Tversky, subjects demanded a 20% reduction in premiums to offset a 1% chance that the insurance company would not pay (Wakker, Thaler, and Tversky 1997). Of course, there is always a chance that an insurance company will refuse to pay. Nonetheless, subjects appear to value the facial certainty provided by insurance contracts; a slight deviation from this facial certainty resulted in a dramatic reduction in the perceived attractiveness of the policy.

One can also use the certainty effect to help explain why government agencies sometimes choose to spend a disproportionate amount of money cleaning up the last vestiges of hazardous waste on, for example, a Superfund site. Stephen Breyer has called this the "last 10% problem." Cleaning up this last 10% often costs far more than the bulk of the cleanup, and results in relatively small benefits. He attributes the last 10% problem to agency tunnel vision whereby agencies get carried away in their single-minded focus on remediation (Breyer 1993). But the certainty effect is also likely to be playing a role here. Agencies, much like the subjects in the above research, might place a disproportionate value on knowing that a site is safe. Of course, no site is entirely free from risk. But when the probability of harm dips below a certain threshold people are likely to treat it like a 0% chance of harm and value such psychological certainty accordingly.

The certainty effect's influence on environmental law potentially extends beyond site cleanups. Cass Sunstein has suggested that the certainty effect can help explain the content of the Clean Air Act. This act requires the EPA to set air quality levels that are "requisite to protect the public health" with "an adequate margin of safety." Sunstein argues that these provisions promise that the air will be "safe," and this idea of perfectly safe air—rather than merely less risky air—was seductive enough that Congress ignored the costs that might be required (Sunstein 2002).

When people value certainty they are not necessarily making an error. Perceived certainty has the unique property of relieving fear and worry. Even a very low-probability risk can trigger fear and worry. Such risks are especially likely to do so if they are emotionally salient. For example, many parents fear that strangers will abduct their children, even though they acknowledge that the risks are vanishingly small (Furedi 2001). Because fear and worry are real psychological costs, reducing or eliminating them is a real benefit that people should be willing to pay something extra for.

Although perceived certainty has benefits, the magnitude of the certainty effect provides strong reasons to question the rationality of many choices. Many people, upon reflection, would choose to allocate resources in ways that would at least reduce the magnitude of the certainty effect. That is, people might prefer to allocate resources to *reducing* large risks like the risk of car accidents rather than *eliminating* less dangerous threats like Alar on apples or an exceedingly small amount of lingering hazardous waste.

## 4 CONCLUSION

---

Overoptimism, ambiguity aversion, and the certainty effect each help explain a plethora of human behavior, and have numerous implications for law. Overoptimism can help explain excessive borrowing, inattention to certain contract terms, underpreparing for divorce, underinsuring against and underpreparing for health risks, and excessive litigation in both civil and criminal contexts. Identifying overoptimism as one source of these negative effects can help policymakers design appropriate legal responses. Ambiguity aversion helps explain patterns in contracting behavior and plea bargaining. It can also be used strategically to increase deterrence and decrease the distortionary effects of the tax code. The certainty effect helps explain the content of environmental regulations and the behavior of environmental agencies. It also suggests that people may be allocating scarce resources in ways that, upon reflection, they would reject. Although translating evidence of these probability errors into specific policy recommendations requires close attention to the particular context at issue and to heterogeneity in the relevant population, the phenomena discussed in this chapter can play an important role in legal reform.

## REFERENCES

- Alicke, M. D., and O. Govorun. 2005. The Better-Than-Average Effect. Pp. 85–106 in *The Self in Social Judgment*, edited by M. Alicke, D. Dunning, and J. I. Krueger. New York: Psychology Press.
- Allais, M. 1953. Le comportement de l'homme rationnel devant le risqué: Critique des postulats et axiomes de l'école Américaine. *Econometrica* 21:503–46.
- Aloysius, J. A. 2005. Ambiguity Aversion and the Equity Premium Puzzle: A Re-examination of Experimental Data on Repeated Gambles. *Journal of Socio-Economics* 34:635–55.
- Armor, D. A., and A. M. Sackett. 2006. Accuracy, Error, and Bias in Predictions for Real versus Hypothetical Events. *Journal of Personality and Social Psychology* 91:583–600.
- Avineri, E., and J. N. Prashker. 2004. Violations of Expected Utility Theory in Route-Choice Stated Preferences: Certainty Effect and Inflation of Small Probabilities. *Transportation Research Record: Journal of the Transportation Research Board* 1894:222–29.
- Aviram, A., and A. Tor. 2003. Overcoming Impediments to Information Sharing. *Alabama Law Review* 55:231–79.
- Babcock, L., and G. Loewenstein. 1997. Explaining Bargaining Impasse: The Role of Self-Serving Biases. *Journal of Economic Perspectives* 11:109–26.
- Baker, L. A., and R. E. Emery. 1993. When Every Relationship Is above Average. *Law and Human Behavior* 17:439–50.
- Baker, T., and P. Siegelman. 2010. Tontines for the Invincibles: Enticing Low Risks into the Health-Insurance Pool with an Idea from Insurance History and Behavioral Economics. *Wisconsin Law Review* 2010:79–120.
- Bar-Gill, O. 2004. Seduction by Plastic. *Northwestern Law Review* 98:1373–434.
- Bar-Gill, O., and E. Warren. 2008. Making Credit Safer. *University of Pennsylvania Law Review* 157:1–101.
- Barber, B. M., and T. Odean. 2001. Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment. *Quarterly Journal of Economics* 116:261–92.
- Baucells, M., and F. H. Heukamp. 2010. Common Ratio Using Delay. *Theory and Decision* 68:149–58.
- Ben-Shahar, O., and J. Pottow. 2006. On the Stickiness of Default Rules. *Florida State University Law Review* 33:651–82.
- Bolger, F., and D. Önköl-Atay. 2004. The Effects of Feedback on Judgmental Interval Predictions. *International Journal of Forecasting* 20:29–39.
- Bossaerts, P., P. Ghirardato, S. Guarnaschelli, and W. R. Zame. 2010. Ambiguity in Asset Markets: Theory and Experiment. *Review of Financial Studies* 23:1325–59.
- Boswell, J. [1791] 1874. *The Life of Samuel Johnson, LL.D.* London: Vicker and Son.
- Breyer, S. G. 1993. *Breaking the Vicious Circle: Toward Effective Risk Regulation*. Cambridge, MA: Harvard University Press.
- Cabantous, L. 2007. Ambiguity Aversion in the Field of Insurance: Insurers' Attitude to Imprecise and Conflicting Probability Estimates. *Theory and Decision* 62:219–40.
- Camerer, C., and D. Lovallo. 1999. Overconfidence and Excess Entry: An Experimental Approach. *American Economic Review* 89:306–18.
- Camerer, C., and M. Weber. 1992. Recent Developments in Modeling Preferences: Uncertainty and Ambiguity. *Journal of Risk and Uncertainty* 5:325–70.
- Chambers, J. R., and P. D. Windschitl. 2004. Biases in Social Comparative Judgments: The Role of Nonmotivated Factors in Above-Average and Comparative-Optimism Effects. *Psychological Bulletin* 130:813–38.



- Chandler, C. C., L. Greening, L. J. Robison, and L. Stoppelbein. 1999. It Can't Happen to Me. . . Or Can It? Conditional Base Rates Affect Subjective Probability Judgments. *Journal of Experimental Psychology: Applied* 5:361–78.
- Daniels, K. C., K. D. Rettig, and R. delMas. 2006. Alternative Formulas for Distributing Parental Incomes at Divorce. *Journal of Family and Economic Issues* 27:4–26.
- Dogan, E., L. Steg, P. Delhomme, and T. Rothengatter. 2011. The Effects of Non-evaluative Feedback on Drivers' Self-Evaluation and Performance. *Accident Analysis and Prevention* 45:522–28.
- Du, N., and D. V. Budescu. 2005. The Effects of Imprecise Probabilities and Outcomes in Evaluating Investment Options. *Management Science* 51:1791–803.
- Dunning, D., C. Heath, and J. M. Suls. 2004. Flawed Self-Assessment Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest* 5:69–106.
- Dunning, D., J. A. Meyerowitz, and A. D. Holzberg. 1989. Ambiguity and Self-Evaluation: The Role of Idiosyncratic Trait Definitions in Self-Serving Assessments of Ability. *Journal of Personality and Social Psychology* 57:1082–90.
- Ellsberg, D. 1961. Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics* 75:643–69.
- Ferraro, P. J. 2010. Know Thyself: Competence and Self-Awareness. *Atlantic Economic Journal* 38:183–96.
- Feldman, Y., and S. Lifshitz. 2011. Behind the Veil of Legal Uncertainty. *Law and Contemporary Problems* 74:133–74.
- Fiske, S. T., and S. E. Taylor. 1984. *Social Cognition*. Reading, MA: Addison-Wesley.
- Fox, C. R., and A. Tversky. 1995. Ambiguity Aversion and Comparative Ignorance. *Quarterly Journal of Economics* 110:585–603.
- Fox, C. R., and M. Weber. 2002. Ambiguity Aversion, Comparative Ignorance, and Decision Context. *Organizational Behavior and Human Decision Processes* 88:476–98.
- Francis, K. E. 2010. Rollover, Rollover: A Behavioral Law and Economics Analysis of the Payday-Loan Industry. *Texas Law Review* 88:611–38.
- Furedi, F. 2001. *Paranoid Parenting*. London: Penguin.
- Gerrard, M., F. X. Gibbons, and M. Reis-Bergan. 1999. The Effect of Risk Communication on Risk Perceptions: The Significance of Individual Differences. *JNCI Monographs* 25:94–100.
- Goodman-Delahunty, J., P. A. Granhag, M. Hartwig, and E. F. Loftus. 2010. Insightful or Wishful: Lawyers' Ability to Predict Case Outcomes. *Psychology, Public Policy, and Law* 16:133–57.
- Halevy, Y. 2007. Ellsberg Revisited: An Experimental Study. *Econometrica* 75:503–36.
- Hayden, B. Y., S. R. Heilbronner, and M. L. Platt. 2010. Ambiguity Aversion in Rhesus Macaques. *Frontiers in Neuroscience* 4:1–7.
- Heller, K. 2006. The Cognitive Psychology of Circumstantial Evidence. *Michigan Law Review* 105:241–305.
- Helzer, E. G., and D. Dunning. 2012. Why and When Peer Prediction Is Superior to Self-Prediction: The Weight Given to Future Aspiration versus Past Achievement. *Journal of Personality and Social Psychology* 103:38–53.
- Ho, J. L. Y., L. R. Keller, and P. Keltyka. 2002. Effects of Outcome and Probabilistic Ambiguity on Managerial Choices. *Journal of Risk and Uncertainty* 24:47–74.
- Hogarth, R. M., and H. J. Einhorn. 1990. Venture Theory: A Model of Decision Weights. *Management Science* 36:780–803.
- Huck, S., and W. Müller. 2012. Allais for All: Revisiting the Paradox in a Large Representative Sample. *Journal of Risk and Uncertainty* 44:261–93.



- Jolls, C., and C. R. Sunstein. 2006. Debiasing through Law. *Journal of Legal Studies* 35:199–242.
- Jolls, C., C. R. Sunstein, and R. Thaler. 1998. A Behavioral Approach to Law and Economics. *Stanford Law Review* 50:1471–550.
- Keller, L. R., R. Sarin, and J. Souderpandian. 2007. An Examination of Ambiguity Aversion: Are Two Heads Better Than One? *Judgment and Decision Making* 2:390–97.
- Keren, G., and L. E. M. Gerritsen. 1999. On the Robustness and Possible Accounts of Ambiguity Aversion. *Acta Psychologica* 103:149–72.
- Kreuter, M. W., and V. J. Strecher. 1995. Changing Inaccurate Perceptions of Health Risk: Results from a Randomized Trial. *Health Psychology* 14:56–63.
- Kunreuther, H., J. Meszaros, R. M. Hogarth, and M. Spranca. 1995. Ambiguity and Underwriter Decision Processes. *Journal of Economic Behavior and Organization* 26:337–52.
- Lawsky, S. B. 2013. Modeling Uncertainty in Tax Law. *Stanford Law Review* 65:241–78.
- Lieberman, N., and Y. Trope. 2008. The Psychology of Transcending the Here and Now. *Science* 322:1201–205.
- Liu, H. H., and A. M. Colman. 2009. Ambiguity Aversion in the Long Run: Repeated Decisions under Risk and Uncertainty. *Journal of Economic Psychology* 3:277–84.
- Loewenstein, G., and D. A. Moore. 2004. When Ignorance is Bliss: Information Exchange and Inefficiency in Bargaining. *Journal of Legal Studies* 33:37–58.
- Mahar, H. 2003. Why Are There So Few Prenuptial Agreements? Harvard Law School John M. Olin Center. for Law, Economics, and Business Discussion Paper Series, Paper no. 436. Available at [http://www.law.harvard.edu/programs/olin\\_center/papers/pdf/436.pdf](http://www.law.harvard.edu/programs/olin_center/papers/pdf/436.pdf).
- Mezulis, A. H., L. Y. Abramson, J. S. Hyde, and B. L. Hankin. 2004. Is There a Universal Positivity Bias in Attributions? A Meta-analytic Review of Individual, Developmental, and Cultural Differences in the Self-Serving Attributional Bias. *Psychological Bulletin* 130:711–47.
- Moore, D. A. 2007. Not So Above Average after All: When People Believe They Are Worse Than Average and Its Implications for Theories of Bias in Social Comparison. *Organizational Behavior and Human Decision Processes* 102:42–58.
- Preston, C. E., and S. Harris. 1965. Psychology of Drivers in Traffic Accidents. *Journal of Applied Psychology* 49:284–88.
- Pulford, B. D. 2009. Is Luck on My Side? Optimism, Pessimism, and Ambiguity Aversion. *Quarterly Journal of Experimental Psychology* 62:1079–87.
- Radcliffe, N. M., and W. M. Klein. 2002. Dispositional, Unrealistic, and Comparative Optimism: Differential Relations with the Knowledge and Processing of Risk Information and Beliefs about Personal Risk. *Personality and Social Psychology Bulletin* 28:836–46.
- Rothman, A. J., and M. T. Kiviniemi. 1999. Treating People with Information: An Analysis and Review of Approaches to Communicating Health Risk Information. *JNCI Monographs* 1999:44–51.
- Rottenstreich, Y., and C. K. Hsee. 2001. Money, Kisses, and Electric Shocks: On the Affective Psychology of Risk. *Psychological Science* 12:185–90.
- Sanitioso, R. B., and R. Wlodarski. 2004. In Search of Information That Confirms a Desired Self-Perception: Motivated Processing of Social Feedback and Choice of Social Interactions. *Personality and Social Psychology Bulletin* 30:412–22.
- Scheier, M. F., and C. S. Carver. 1985. Optimism, Coping, and Health: Assessment and Implications of Generalized Outcome Expectancies. *Health Psychology* 4:219–47.
- Sedikides, C., and M. D. Alicke. 2012. Self-Enhancement and Self-Protection Motives. Pp. 303–22 in *Oxford Handbook of Motivation*, edited by R. M. Ryan. New York: Oxford University Press.

- Sedikides, C., and J. D. Green. 2004. What I Don't Recall Can't Hurt Me: Information Negativity versus Information Inconsistency as Determinants of Memorial Self-Defense. *Social Cognition* 22:4–29.
- Sedikides, C., and A. P. Gregg. 2008. Self-Enhancement: Food for Thought. *Perspectives on Psychological Science* 3:102–16.
- Segal, U., and A. Stein. 2006. Ambiguity Aversion and the Criminal Process. *Notre Dame Law Review* 81:1495–552.
- Simons, D. J. 2013. Unskilled and Optimistic: Overconfident Predictions Despite Calibrated Knowledge of Relative Skill. *Psychonomic Bulletin and Review* 20:601–607
- Srivastava, S., K. M. McGonigal, J. M. Richards, E. A. Butler, and J. J. Gross. 2006. Optimism in Close Relationships: How Seeing Things in a Positive Light Makes Them So. *Journal of Personality and Social Psychology* 91:143–53.
- Sunstein, C. R. 2003. Terrorism and Probability Neglect. *Journal of Risk and Uncertainty* 26:121–36.
- Sunstein, C. R. 2002. *The Cost-Benefit State: The Future of Regulatory Protection*. Chicago: American Bar Association.
- Talley, E. L. 2009. On Uncertainty, Ambiguity, and Contractual Conditions. *Delaware Journal of Corporate Law* 34:755–812.
- Thompson, L., and G. Loewenstein. 1992. Egocentric Interpretations of Fairness and Interpersonal Conflict. *Organizational Behavior and Human Decision Processes* 51:176–97.
- Trautmann, S. T., F. M. Vieider, and P. P. Wakker. 2008. Causes of Ambiguity Aversion: Known versus Unknown Preferences. *Journal of Risk and Uncertainty* 36:225–43.
- Tversky, A., and D. Kahneman. 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* 5:297–323.
- Viscusi, W. K., and H. Chesson. 1999. Hopes and Fears: The Conflicting Effects of Risk Ambiguity. *Theory and Decision* 47:157–84.
- Wakker, P., R. Thaler, and A. Tversky. 1997. Probabilistic Insurance. *Journal of Risk and Uncertainty* 15:7–28.
- Weber, B. J., and G. B. Chapman. 2005. The Combined Effects of Risk and Time on Choice: Does Uncertainty Eliminate the Immediacy Effect? Does Delay Eliminate the Certainty Effect? *Organizational Behavior and Human Decision Processes* 96:104–18.
- Weinstein, N. D. 1980. Unrealistic Optimism about Future Life Events. *Journal of Personality and Social Psychology* 39:806–20.
- Weinstein, N. D. 1987. Unrealistic Optimism about Susceptibility to Health Problems: Conclusions from a Community-Wide Sample. *Journal of Behavioral Medicine* 10:481–500.
- Weinstein, N. D., and W. M. Klein. 1995. Resistance of Personal Risk Perceptions to Debiasing Interventions. *Health Psychology* 14:132–40.
- Williams, S. H. 2008. Sticky Expectations: Responses to Persistent Over-optimism in Marriage, Employment Contracts, and Credit Card Use. *Notre Dame Law Review* 84:733–91.

## CHAPTER 14

---

# THE HINDSIGHT BIAS AND THE LAW IN HINDSIGHT

---

DORON TEICHMAN

### 1 INTRODUCTION

---

WE are all Monday-morning quarterbacks; that is the key finding of the huge body of literature documenting the hindsight bias. When asked to judge events in retrospect, we find it difficult to screen outcome information that we were exposed to. This phenomenon stems from both cognitive forces—it is difficult to train our mind to ignore information when we constantly train it to incorporate all available information—as well as motivational factors—it is more pleasant to perceive ourselves as those who were sharp enough to predict the unpredictable.

Alongside its influence on the psychological literature, the hindsight bias has had a profound effect on legal scholarship as well. A Westlaw search in the *Journals and Law Review* database reveals 1,293 documents alluding to the precise term “hindsight bias” from 1982 till the present (this count arguably reflects some false positives and many false negatives). Baruch Fischhoff, the psychologist who first identified the bias and who is most associated with it, is cited in 412 articles. And Rachlinski (1998), the article that offers the first comprehensive legal analysis of the bias, has been cited since its publication 256 times.

This chapter reviews this vast body of work. It highlights the different legal fields in which the bias is of relevance, and presents the main empirical findings that demonstrate how it actually influences legal decision-making. In addition, the chapter will assess the normative implications of the bias, and describe some of the main legal policies proposed in order to deal with it.

## 2 THE HINDSIGHT BIAS: EMPIRICAL AND THEORETICAL BACKGROUND

---

The hindsight bias—or, the “Monday-morning quarterback” bias, or the “I knew it all along” bias—refers to the tendency of people to overestimate the probability of an event once they are aware of the fact that the event has occurred. The initial contribution to the study of this bias was presented by Fischhoff (1975). In this classic study Fischhoff asked his subjects to read a detailed description of the historical background leading to the nineteenth-century British-Gurka war, and then requested them to estimate the likelihood of four different potential outcomes of the event. Unbeknownst to the participants they were randomly assigned to either a foresight or a hindsight condition. Participants in the foresight group were given no outcome information. Participants in the hindsight groups, on the other hand, were informed that one of the potential outcomes was the “true” outcome of the event. The results of the experiment showed that subjects were unable to ignore outcome information. Once participants were told that a certain outcome occurred, they tended to view it as significantly more likely to happen.

The basic result of Fischhoff (1975) has been replicated in dozens of studies. These studies changed the initial experimental setup along dimensions such as methodology (e.g., between subject vs. within subject), type of population (e.g., laypeople vs. experts), decision-making environment (e.g., individual vs. group); and the context in which predictions were made (e.g., hypothetical historical stories vs. professional case studies). These experiments have demonstrated the wide scope and robustness of the bias. A comprehensive review and meta-analysis of the hindsight literature can be found in Hawkins and Hastie (1990), Christensen-Szalanski and Willham (1991), Guilbault and coauthors (2004) and Roese and Vohs (2012).

While the existence of the bias is undisputed, researchers have highlighted distinct underlying processes that might explain it (Guilbault et al. 2004). The first group of explanations focuses on the cognitive aspects of the bias. According to this perspective people search their memories for old beliefs that are confirmed by the outcome information. Alternatively, outcome information might be an anchor point that serves as a starting point for the probability estimation process. A second cluster of explanations focuses on motivational aspects. According to this line of thought people want to perceive themselves in a favorable light. Since the ability to predict events precisely is noteworthy, people tend to overstate their ability to do so.

Finally, it should be noted that Baron and Hershey (1988) document a highly related bias—the outcome bias. Whereas the central point in the hindsight bias is the *ex post* adjustment of probability assessments, the outcome bias focuses on the question whether outcomes influence the way in which people judge the quality of a decision. As Baron and Hershey (1988) show, even when the *ex ante* probabilities are clear and

transparent (e.g., monetary bets with explicit odds), people tend to judge choices that led to unfavorable outcomes more harshly than choices that led to favorable outcomes. This phenomenon has been replicated in other settings as well, in both expert and non-expert populations (Caplan, Posner, and Cheney 1991; Gino, Shu, and Bazerman 2010). For the most part, this chapter will focus on the role of the hindsight bias in legal analysis, though its arguments are tied to the outcome bias as well.

### 3 HINDSIGHT BIAS IN LEGAL DECISION-MAKING: EVIDENCE

---

By its very nature the legal process often focuses on an ex post evaluation of people's conduct. In such cases judges and jurors are required to ignore outcome information and assess decisions given the information set that was available during the time they were made. To the extent judges and jurors are subject to the hindsight bias, they are expected to systematically attribute responsibility to decisions that were made in accordance with the legal standard when viewed from an ex ante perspective.

Early legal scholarship relied on psychological studies that documented the bias, and theorized that it extends to legal decision-making as well (Loftus and Beach 1982). As the field of empirical legal studies grew, an independent body of work demonstrated the prevalence of the bias in legal decision-making as well. These studies reveal that decisions relating to both substantive and procedural questions are influenced by the bias. Following is a brief review of the main findings in this body of work.

#### 3.1 Tort Law

A natural place to begin a review of the hindsight bias in legal decision-making is tort law, and more specifically negligence law. Negligence cases involve situations in which a plaintiff asserts that she was harmed due to the fact that the defendant failed to exercise a reasonable level of care. Theoretically, the focus of such cases should be on the decision made by the defendant with respect to precautions at the time in which it was made (*Restatement (Third) of Torts: Liability for Physical Harm*, §3). Thus, in such cases legal decision-makers are expected to ignore the fact that an accident occurred when determining whether the defendant took due care.

In what has become a classic study Kamin and Rachlinski (1995) demonstrated that ex ante evaluation of precautions differ significantly from ex post evaluations of the same precautions. Participants in the study were asked to evaluate a city's decision regarding the proper precautions that should be taken in order to deal with the risk of a flood. All subjects received the same facts regarding the costs and effectiveness of the precaution. They differed, however, with respect to the perspective in which they

evaluated the precaution. Whereas subjects in the foresight group simulated an administrative hearing that was required to make decisions with respect to precautions prior to any harm materializing, subjects in the hindsight group simulated a trial that was held after a harm had materialized. The results showed a significant difference between the two groups: participants in the hindsight group were far more likely to determine that the precaution should be taken. Similar results were documented by LaBine and LaBine (1996) in the context of the precautions needed to be taken by a therapist in light of a risk of a patient behaving violently. Their findings show that holding everything else equal, participants were more likely to rule the therapist negligent in cases in which they were informed that harm materialized.

Other studies have attempted to document the bias in the context of actual decisions made in the courtroom. Cheney and his colleagues (1989) examined 1,004 court cases alleging anesthesia-related negligence and found that in over 40% of the cases in which liability was found the physician acted appropriately. Using a similar methodology Taragin and his colleagues (1992) showed that in 21% of the cases that they examined physicians were found liable for defensible practices.

### 3.2 Corporate Law

A similar legal context that requires an *ex post* evaluation of care under a reasonableness standard can be found in corporate litigation. When investors claim that corporate officers made negligent business decisions they in effect ask courts to re-examine those decisions while ignoring their outcomes. As is the case in regular negligence cases, applying a reasonableness test to boardroom decisions after those decisions generated disappointing results raises concern that biased courts might determine that reasonable (yet risky) decisions were unreasonable.

In order to examine this hypothesis Stallard and Worthington (1998) designed a controlled experiment. All subjects in the experiment were presented with materials that were based on the facts of a real case that dealt with the responsibility of board members for the failure of the corporation. Participants in the hindsight group were informed of the outcome of the board's decision and were asked to evaluate it as part of an *ex post* adjudicative process. Participants in the foresight group, on the other hand, were not informed of the outcome, and were asked to evaluate the board's decision as part of a review hearing. The results revealed a significant difference between the two groups: members of the hindsight group were more likely to determine that the board acted negligently.

### 3.3 Patent Law

The hindsight bias is not relevant exclusively to the assignment of legal liability in cases in which harms materialize. Researchers have demonstrated that patent litigation can also be influenced by the bias. For a patent to be valid its holder must demonstrate that

the invention was “non-obvious” (35 U.S.C. §§ 101-03, 112). Thus, much like the foregoing examples, patent jurisprudence also requires legal decision-makers to travel back in time. In the patent context the destination of this travel is the preinvention period. Only once legal decision-makers situated themselves in that time frame should they ask themselves whether the invention represents a nontrivial step forward. Conducting this analysis in hindsight, however, might be a tricky task since it requires people to ignore the fact that the invention was achieved. Decision-makers influenced by the bias are expected to systematically view nonobvious inventions as obvious. The concern over the bias has even manifested in court rulings that have cautioned against “slipping into use of hindsight” when determining what is obvious and what is not (*Graham v. John Deere Co.* 1966, 36).

Using controlled experiments Mandel (2006; 2006–7) examined the effect of hindsight on the determination of nonobviousness. In his studies participants were asked to evaluate whether developing a solution to a certain technological challenge constituted a nonobvious development. Participants in the foresight condition were presented with the background information alone, while participants in the hindsight condition were presented with the same background information and with the actual solution that the inventor had thought of. The results confirmed that the hindsight bias influences the evaluation of patents. Participants were far more likely to find the invention to be obvious in the hindsight condition than they were in the foresight condition.

### 3.4 Criminal Procedure

Thus far we have seen that the hindsight bias might greatly influence the way in which substantive legal rules are applied. Additional studies have demonstrated that the bias might also take center stage in procedural contexts. In cases involving police searches courts are often required to determine whether the officers who conducted them had a probable cause to do so. This determination, however, might arise in both foresight and hindsight situations. Foresight situations involve police officers who wish to obtain a warrant prior to conducting the search. In such cases judges are required to rule on the legality of the search without knowing its outcome. Hindsight situations involve evidentiary rulings regarding the admissibility of evidence obtained in the search. Such cases by definition only involve situations in which the police officer found incriminating evidence during the search. The hindsight bias suggests that judges will exhibit a greater tendency to rule in the latter category of cases that the search was based on a probable cause since their judgment of the search will be influenced by the fact that in retrospect it turned out to be justified.

This hypothesis was tested extensively in a series of neatly designed experiments that were run on several hundred judges in the United States (Wistrich, Guthrie, and Rachlinski 2005; Guthrie, Rachlinski, and Wistrich 2007; Rachlinski, Guthrie, and Wistrich 2011). The framework of all of the experiments reported in these papers was similar. Judges read a scenario describing a search conducted by a policeman. The



scenarios were intentionally ambiguous and raised doubt as to the existence of a probable cause. The experimental manipulation was created by the legal context in which the scenario was embedded. Participants in the foresight group evaluated the situation as if they were asked to grant a search warrant, while participants in the hindsight group evaluated it for purposes of admissibility. Interestingly, the general picture arising from this wide body of research is that judges did not differ significantly between the foresight and the hindsight conditions.<sup>1</sup>

### 3.5 Legal Experts

Numerous studies have explored whether experts outperform laypeople with respect to the hindsight bias, and the findings have been mixed. Whereas a meta-analysis by Christensen-Szalanski and Willham (1991) suggests that experts are less influenced by the bias, a subsequent meta-analysis by Guilbault and coauthors (2004) found no such relationship. This result might stem from the fact that different experts make distinct decisions. Whereas experts such as doctors and weather forecasters constantly receive feedback as to the accuracy of their predictions, stockbrokers and clinical psychologists do not receive such clear feedback. In light of these ambiguities, and given that in some legal systems decision-making is vested exclusively in the hands of judges (and that in jury-based systems judges still rule in many of the cases), documenting the performance of judges is of specific importance.

Guthrie, Rachlinski, and Wistrich (2000) examined the hindsight bias in an experiment they conducted with a group of judges. Subjects in the experiment read the description of a real case, and were then asked to predict how the appellate court would rule on it. Participants were randomly assigned into one of three groups that were provided with information as to how the appellate court ruled in the case (i.e., affirmed, vacated, lesser sanction). The information regarding the outcome influenced the judges' predictions significantly. For example, whereas in the Affirmed group 81.5% of the judges believed that the appellate court would affirm, only 7.4% and 11.1% believed so in the Lesser Sanction and Vacated groups (respectively). As the authors note, this suggests that "[l]earning the alleged outcome on appeal influenced the judges' assessments of what the likely outcome would have been" (2000, 803).

While Guthrie, Rachlinski, and Wistrich (2000) documented hindsight bias among judges, one should acknowledge the study's limitation as well. Judges in this study faced a somewhat synthetic task of predicting the result of an appellate review. A more direct evaluation of the relative ability of judges to deal with the bias was presented in Hastie

<sup>1</sup> In a couple of related papers Casper, Benedict, and Kelly (1988, 1989) demonstrated that search-outcome information affected the level of damages awarded by mock jurors in cases involving allegedly illegal searches. These studies do not fully document the effect of the hindsight bias as it is quite plausible that the outcome of the search influenced subjects' judgment of the plaintiff and not of the policemen.

and Viscusi (1998) (see also Anderson et al. 1995; Viscusi 1999). The study employed a tort-precaution framework similar to those in other studies. Importantly, the population of subjects included 95 judges and 277 mock jurors that faced identical survey instruments. This design feature enabled the study to measure the difference between the two populations. The results suggest that the two groups were distinct. While mock jurors exhibited a clear hindsight bias, judges only exhibited trends towards hindsight (that were mostly statistically insignificant). Based on these findings the authors concluded that “judges exhibit much smaller hindsight effects when asked to make identical judgments” (1998, 917).

\* \* \*

The general picture arising from this section is that the hindsight bias influences legal decisions in a significant manner. When adjudication requires an ex post evaluation of behavior, courts might fail to ignore outcome information. That said, however, as the last two subsections demonstrate, while judges are not immune from the bias, there are indications that its effect on their behavior is significantly smaller than its effect on untrained individuals.

## 4 THE IMPLICATIONS OF JUDGING IN HINDSIGHT

---

This section analyzes the normative implications of the hindsight bias for legal policy-making. First, it introduces the concerns raised with respect to ex post judicial decisions that are highly susceptible to the bias. This analysis will focus exclusively on the hindsight bias, and will highlight the potential problems associated with judging in hindsight. That done, I will introduce into the analysis the complexities associated with other cognitive biases. Once these additional factors are accounted for, the normative picture becomes somewhat more nuanced.

A methodological note is necessary at the outset. The analysis in this section will be from a purely consequentialist perspective. Thus, it will focus on the inefficiencies that might be created by judgments that are skewed by the bias. That said, however, the bias might also have normative ramifications from nonconsequentialist points of view. It could be unfair to judge people ex post according to a standard it would be very difficult to live up to ex ante. This type of analysis lies outside the scope of this section (and this chapter).

### 4.1 Judging in Hindsight: The Effects of the Hindsight Bias

The hindsight bias might generate erroneous legal decisions. If judgments made during adjudication cannot ignore outcome information, then those judgments are expected to be systematically skewed. Adjudicators making decisions ex post will tend to assume

that certain events were more likely than the regulated agent could have anticipated *ex ante*. As a result, a biased judicial system is likely to be misaligned with people's *ex ante* assessments of the situation subject to legal regulation.

The key concern often raised with respect to the bias is that the judgments rendered by judicial decision-makers *ex post* will generate perverse incentives for agents making decisions *ex ante*. In the context of negligence law, for example, fact-finders functioning in hindsight might systematically overestimate the probability of harmful events. As a result, they are expected to broaden the scope of negligence liability inefficiently. Potential tortfeasors who are aware of this type of judging that awaits them down the road, are expected to react to it by taking excessive care.

A more careful analysis, however, suggests that the precise ramifications of the bias are more ambiguous (Rachlinski 1998). On one hand, if the bias creates a large deviation from the optimal level of care, then this implies that a biased-negligence regime operates as a *de facto* strict-liability regime. As conventional economic analysis has taught us, potential tortfeasors facing a strict liability regime are not expected to take excessive care since their cost-reducing strategy is to take optimal care (Shavell 1985). On the other hand, if the effect of the bias on the determination of the negligence standard is moderate, then potential tortfeasors might be driven to take excessive precautions and meet this inefficiently high standard. This is because the immunity from tort liability for nonnegligent harms granted by a negligence regime will justify the moderate costs of additional inefficient care. Regretfully, as Rachlinski (1998, 599) notes, “[n]either common law judges nor legislators are likely to have the capacity to distinguish which of these two situations any class of cases follows.” All of this suggests that while a biased-negligence regime is not optimal, its precise implications are unclear.

Similar attention should be given to other areas of law before any definitive normative conclusions can be drawn. In the context of nonobviousness litigation in patent law, for example, it is unclear whether the biased findings of obviousness are undesirable. To the extent patent law offers inventors excessive protection that creates unjustifiable monopolistic costs (Burk and Lemley 2009), biased rulings that systematically function against inventors might arguably help maximize social welfare as a means through which the law narrows the scope of patent protection.

Thus far the analysis has focused on the effect of biased judgments on the primary behavior (i.e., how much care people take, how many inventions they develop). An additional way in which the bias might influence behavior is through the incentives it creates with respect to the production of evidence. Potential litigants are aware of the difficulties associated with proving disputed facts to a detached third party who is limited to basing her decision on observable and verifiable evidence. As Stein and Parchomovsky (2010) note, “rational actors will always interpret the dictates of our substantive law through an evidentiary gloss, which in many cases will prompt actors to deviate from the outcome envisioned by efficiency-minded legislatures and courts” (521). Hence, such actors might choose *ex ante* to behave inefficiently in order to produce a sufficient amount of evidence that will enable them to prevail in *ex post* litigation.

In the context of the hindsight bias, potential litigants might spend significant resources in order to build an evidentiary picture that will enable them to haul the judicial fact-finder back in time and diminish the effect of the bias. For example, a board might pay for an extensive expert opinion that explains the risks associated with a certain decision, even if all of those risks are perfectly clear to all parties involved. The costs associated with the production of such superfluous evidence should also be seen as part of the inefficiencies generated by the bias.

## 4.2 Judging in Hindsight: Countervailing Factors

So far the analysis has focused on the hindsight bias in isolation. A more complex picture emerges when we incorporate into the discussion other biases that might affect the ultimate design of the liability regime. Of specific relevance in the context of tort liability is the tendency of people to be overly optimistic about their abilities and future prospects (see generally chapter 13 by Williams in this volume). As a result of this tendency, people might take insufficient precautions since they assume that the probability that those precautions will be needed is smaller than it actually is. Arguably, a biased negligence regime that somewhat drives people to take more precautions might serve as an imperfect countervailing force that helps incentivize overly optimistic people to take efficient precautions.

Jolls, Sunstein, and Thaler (1998, 1524–25) are dismissive of this possibility. After assuming that this bias is of diminished importance with respect to corporations (a disputed assumption that should be examined: see chapter 21 by Tor in this volume), they move on to note that for individuals “the role of over optimism is likely to vary significantly with context,” an uncontroversial observation. They then continue to suggest that the salience of tort cases is actually expected to cause potential tortfeasors to overestimate the probability of liability. Given these ambiguous arguments and the “across the board” nature of the hindsight bias Jolls, Sunstein, and Thaler conclude that overoptimism does not affect the analysis.

This conclusion seems difficult to defend. While the hindsight bias is an “across the board” phenomenon, overoptimism is also a robust pattern of behavior, which has been documented in an array of contexts. Furthermore, it is far from clear whether salience plays any significant role in the context of tort law. For most people (i.e., people who are not tort professors), tort law is not salient or available. If anything, the routine in which people constantly engage in risky activities (e.g., driving) without harm materializing suggests that they will become less sensitive to the legal system (Hertwig and Erev 2009). Interestingly, when Jolls analyzes redistributive legal rules, she seems to acknowledge this point and notes that “[i]t is difficult to come up with examples of events giving rise to individual liability the probability of which is likely to be overestimated rather than. . . underestimated” (1998, 1663). Thus, it is quite possible that the hindsight bias could help correct people’s biased decisions with respect to precautions.

Similarly, the analysis of patent law might change once we introduce overoptimism. If patent law lays down optimal rules, and if innovators anticipate that these perfect rules will be applied in a biased fashion, then there is reason to worry that the incentives set by the legal system will be inefficient. If, however, innovators exhibit irrational optimism regarding their ability to develop patentable technologies, then biased adjudication might be desirable. Overly optimistic innovators will continue to develop useful technologies believing they will enjoy patent protection. After the fact, biased courts will not grant these technologies patent protection, and thus the deadweight losses associated with such protection will be avoided. If you will, society will get to have its cake and eat it too.

A different behavioral phenomenon that is of relevance to the evaluation of the hindsight bias is motivated reasoning. According to this body of literature, decision-makers attempt to make choices that they perceive as such that can be later justified to a dispassionate observer. In other words, people want to avoid a dissonance between how they behave and how they think they ought to behave. Thus, they develop an array of internal reasoning tools that allow them to justify their acts, and avoid categorizing them as immoral or dishonest (for a review see chapter 9 by Feldman in this volume).

In the legal context motivated reasoning suggests that people might interpret vague standards such as “negligence” in a self-serving manner. Feldman and Teichman (2009) explored this possibility by documenting the way in which enforcement uncertainty (e.g., will I get caught?) and legal uncertainty (e.g., will a judge find me negligent?) influence people’s willingness to engage in antisocial behavior. They found that holding the size of the expected sanction constant, legal uncertainty generates a higher level of noncompliance. From this finding Feldman and Teichman deduced that “[e]x ante legal uncertainty allows people to justify their choices to themselves by focusing on the possibility that their acts may be deemed legal” (2009, 1010).

To the extent motivated reasoning influences the way in which people interpret vague standards such as negligence, this would suggest that their subjective ex ante interpretation is significantly less demanding than the objective standard. The hindsight bias might help overcome this problem, by elevating the ultimate standard people are required to live up to. Again, the ex post mistake made in the courtroom helps counterbalance an ex ante mistake made by potential tortfeasors.

\* \* \*

The hindsight bias raises a concern that misalignment between ex ante expectations and ex post adjudication will bring about undesirable results. That said, however, as the analysis in this section suggests, any normative claim regarding the bias requires dealing with two additional factors. First, is the substantive rule at hand structured such that the application of the bias will have a beneficial (or null) effect? Second, are there countervailing external forces that impede efficiency that the bias can help overcome? The fact that courts are getting cases “wrong” because of the bias does not necessarily imply that this is a bad outcome.

## 5 THE HINDSIGHT BIAS AND THE LAW

---

Notwithstanding the above discussion, this section assumes that the hindsight bias creates social costs, and therefore the legal system should attempt to deal with it. As will be shown, legal systems employ both procedural tools that aim to eliminate the effect of the bias on decision-makers and substantive tools that shift the focal point of legal disputes from *ex post* evaluations of behavior to *ex ante* assessments. While none of these measures offers a complete solution to the problems associated with the bias, they do alleviate some of them.

### 5.1 Debiasing

A logical starting point for the discussion on legal reactions to the bias is the possibility of debiasing decision-makers. After all, if the behavior of decision-makers is the source of the problem, regulating it sounds like a sensible solution. Despite its logical appeal, however, this line of thought has been mostly unconstructive. As the following review suggests, a significant body of work has demonstrated that undoing the effects of the hindsight bias is a thorny task.

Soon after the publication of Fischhoff (1975) researchers turned to explore whether certain interventions can reduce the effect of the hindsight bias. The general picture arising from this body of work does not reflect great success. Merely warning people about the bias does not seem to help (Fischhoff 1977); neither does offering them monetary incentives (Hell et al. 1988). A meta-analysis of close to 100 hindsight studies concluded that “manipulations to reduce hindsight bias did not result in significantly smaller effect sizes. . . than studies in which no manipulations to increase or reduce hindsight bias were included” (Guilbault et al. 2004, 111).

Despite this general picture, one debiasing approach that has proven relatively effective is the consider-the-opposite strategy. According to this strategy decision-makers are encouraged to actively think about counterfactual scenarios that do not involve the outcome that materialized (Lord, Lepper, and Preston 1984; Nario and Branscombe 1995; Tetlock 2005). Dawson and colleagues (1988) demonstrated the effectiveness of the strategy in the context of the evaluation of medical decisions. In the study physicians were asked to estimate the probability of three different diagnoses given a certain clinical picture. Participants in the foresight group were not informed of the “true” diagnosis, whereas participants in the hindsight group were informed that one of the three diagnoses was the one that eventually turned out to be correct. The hindsight-debias group was designed like the hindsight group, aside from the fact that its participants were asked to list one reason why each of the diagnoses might be correct. The results showed that this procedure brought about a significant decline in the size of the bias.

While participants in the hindsight-debias group were still influenced by the exposure to outcome information, this influence was lower than in the hindsight group.

To be sure, one should be cautious as to the amount of effort decision-makers are asked to put into identifying alternative outcomes. Sanna, Schwarz, and Stocker (2002) and Sanna and Schwarz (2003) applied two debiasing procedures to their subjects. Whereas half of them were asked to think of a small number of alternative outcomes (e.g., two) the other half was asked to think of a large number of alternative outcomes (e.g., ten). Interestingly, the second debiasing technique backfired and exacerbated the magnitude of the bias. The authors hypothesized that since coming up with a large number of alternative outcomes was relatively difficult, participants concluded from this difficulty that the outcome they were informed about was more likely.

A few studies have explored the effectiveness of concrete debiasing measures in legal settings. In Kamin and Rachlinski (1995, 97) participants were informed by the judge:

Making a fair determination of probability may be difficult. As we all know, hindsight vision is always 20/20. Therefore it is extremely important that before you determine the probability of the outcome that did occur, you fully explore all other possible alternative outcomes which could have occurred. Please take a moment to think of all the ways in which the event in question may have happened differently or not at all.

This procedure did not render any meaningful difference between the debiased group and the group that did not undergo such a procedure. That said, however, one should note that this procedure is much closer to an admonishment procedure. Effective consider-the-opposite strategies require forcing decision-makers to actually consider the opposite by, for example, asking them to write down alternative scenarios (Dawson et al. 1988).

A more aggressive and effective debiasing approach was presented by Lowe and Reckers (1994). Their experiment involved a hypothetical lawsuit brought against an accounting firm that audited a company and issued a favorable report. Participants who were required to evaluate the audit report were either given no outcome information or were informed that the company eventually filed for bankruptcy. Participants in the hindsight-debias group went through a three-stage debiasing procedure. They were presented with two alternative scenarios in which the company remained solvent; they were asked to assess the likelihood of those scenarios; and, finally, they were required to come up with their own alternative scenario. The debiasing procedure proved effective: participants in this group evaluated the auditors' decisions significantly more favorably than those in the hindsight group with no debiasing procedure.

A final and separate question is whether lawyers can debias decision-makers. Stallard and Worthington (1998) presented results that suggest that this question can be answered (somewhat) affirmatively. In their study they did not rely on the bland instructions presented by a judge. Rather, they presented mock jurors with more vivid



arguments made by the defendant's attorney. This lawyer warned the mock jurors of "Monday morning quarterbacking" and urged them not to second-guess the decisions in hindsight. This quite realistic and simple debiasing technique managed to reduce the bias significantly. Whereas 57% of participants reached a negligence conclusion in the regular hindsight group, this figure dropped to 37% in the debiased group. While this final figure is still higher than the amount of negligence findings in the foresight group (that was only 29%), it does reflect a meaningful reduction in the bias that was achieved by using a very simple tool. More research—conducted both in the lab and in the courtroom—could shed additional light on this important question.

In sum, legal scholars have tended to view debiasing procedure as a nonviable solution to the effects of the bias (Rachlinski 1998). To the extent this argument refers to the possibility of completely eradicating the influence of the bias on decision-makers, it is undoubtedly correct. However, as this review suggests, reducing the effect of the bias through different debiasing techniques is possible. Additional research is required in concrete legal settings to map the contexts in which such strategies can be used, and what is the size of the effect they manage to achieve.

## 5.2 Bifurcation

To the extent debiasing decision-makers proves an unfruitful path, policymakers may consider alternative options that will aim to shield the legal process from the effects of the bias. In this respect scholars have proposed to bifurcate tort proceeding such that the negligence question will be dealt with separately (Poythress 1989; Jolls, Sunstein, and Thaler 1998). For example, Jolls, Sunstein, and Thaler propose that in a medical malpractice case involving a balancing between two risky options (e.g., cesarean section vs. vaginal birth), jurors will be presented with the *ex ante* facts that the physician held, and will not be informed of the outcome of the procedure. Once jurors are unaware of the outcome, so the argument goes, the problems of judging in hindsight can be eliminated.

While bifurcation might prove to be useful in cases involving choices between two risky strategies, it is far less useful in most tort cases. As Korobkin and Ulen (2000) pointed out, in the majority of such cases jurors are asked to examine whether the tortfeasor was required to take additional (costly) safety measures. In other words, the judgment is not between measure X and measure Y, but between adopting a level of care of X or 2X. Within this paradigm there is no way to insulate fact-finders from the occurrence of a harm, as they are well aware of the fact that a harm is an underlying fact of the litigation.

Interestingly, the case for a general bifurcation policy might stem from the outcome bias rather than the hindsight bias. As noted earlier, this bias suggests that the size of the harm itself might alter people's judgment of the tortfeasor's behavior. By screening away outcome information, the effect of this bias might be diminished, and court rulings will better reflect a judgment of the decisions people made. That said, however, a

key caveat should be highlighted. Since the occurrence of a harm cannot be hid from fact-finders in tort cases, they might develop estimates of the magnitude of this harm. It is unclear if and in what way such estimates will influence decisions.

### 5.3 Burden of Proof

An alternative method proposed in the literature to deal with the hindsight bias relates to the burden of proof (Jolls, Sunstein, and Thaler 1998). The traditional evidentiary standard in civil cases is preponderance of the evidence. If the jurors believe that it is more likely than not that the defendant's negligence caused the defendant's harm, they should accept the claim. In a first-best world this evidentiary standard is expected to minimize the number of errors associated with adjudication (Vars 2010). If, however, fact-finders are systematically biased in favor of plaintiffs then this analysis no longer holds, and it might be desirable to raise the evidentiary threshold in order to counteract the bias. Jolls and her colleagues conjecture that raising the evidentiary threshold in civil cases involving hindsight judgments to a level of 75% might be desirable.

Recent findings in the area of judicial decision-making suggest that there might be no need to reform evidentiary thresholds since the de facto meaning of the preponderance of the evidence standard is significantly higher. Zamir and Ritov (2012) examine the connection between the omission bias and the burden of proof. They hypothesize that since dismissal of a claim is likely to be viewed by the judicial decision-makers as an omission while acceptance of a claim is likely to be viewed as a commission, adjudicators will set a higher threshold for the acceptance of claims. To test this hypothesis they design a series of neatly structured experiments that elicit from participants their decision threshold. Their results suggest that the actual threshold is significantly higher than that of the preponderance of the evidence, and lies somewhere between .60 and .75.

### 5.4 Substantive Legal Rules

An alternative to regulating the judicial decision-making process might be altering substantive legal rules such that their application will overcome (or, alas, diminish) the costs associated with the bias. The hindsight bias comes into play when adjudicators need to evaluate choices *ex post*. Numerous legal policies can substitute this structure of regulation with alternative regimes that shift the analysis towards benchmarks that were created *ex ante*. By shifting the analysis in such a fashion some of the problems created by the bias can be alleviated.

One direction the law might take is towards the use of standards that are determined in advance. Since such standards are set prior to the outcome occurring, they might reflect an unbiased assessment of risks. The two main tools that come to mind in this regard are regulations set by government agencies and customs created by the relevant

industry. A rule that views compliance with such standards as a complete defense to negligence liability would negate the problems associated with ex post evaluation (Rachlinski 1998).

In practice, however, courts refuse for the most part to adopt a deferential stance towards compliance with regulations or customs. While compliance with such standards can be introduced as evidence of nonnegligence, it does not offer a complete defense (*Restatement (Third) of Torts: Liability for Physical and Emotional Harm*, §13(a)). This legal policy is grounded in concerns over issues such as sustaining incentives to innovate precautions (Parchomovsky and Stein 2008), and the possibility that government agencies are captured by interest groups (Bernstein 1955). Furthermore, secondary considerations such as enforcement costs also suggest that this policy might be justified (Shavell 2012). The outcome of this policy, however, is that potential tortfeasors cannot avoid the risks associated with hindsight judgments by relying on compliance with customs and regulations.

Despite the general trend manifested in the *Restatement of Torts*, there are concrete contexts in which the law does defer to standards set ex ante. Certain “safe harbors” enable tortfeasors that take the designated precautions to preempt future liability. A broad area of tort law ranging from product liability (Klass 2009) to dram shop liability (Turner 2009) is governed by such safe harbors. In the area of medical malpractice there are many calling for the creation of similar safe harbors based on evidence-based standards (Blumstein 2006).

A different, and to some degree diametrically opposite, strategy the law can adopt towards the bias is to adopt a no-liability regime (Rachlinski 1998). In such cases the law recognizes that given the low quality of decisions produced by adjudicators, it is best it leave the area free from formal legal regulation. Thus, the law creates hurdles to litigation that preempt the filing of lawsuits.

Probably the most famous example of this line of action can be found in the area of corporate law and the application of the Business Judgment Rule (BJR). Generally speaking, the BJR represents nearly complete deference to the decisions of corporate officers and directors with respect to questions relating to the duty of care (i.e., negligent management). According to the rule, as long as the officer or director made an informed decision and was not an interested party, she will not be held liable for her decisions, even if they turned out to be terribly bad in retrospect (*The Walt Disney Co. Derivative Litigation* 2006, 52).

The BJR reflects an understanding that biased liability in the corporate context might be damaging. Corporate decisions by their very nature involve trade-offs between risk and returns. Shareholders want managers to take efficient risks, knowing that in some cases those risks will turn out to be losing bets, since they can hedge those risks through prudent diversification. While in a perfect world it could have been beneficial to have a neutral entity examine whether the choices made by management were efficient, this is not the case in a world with biased courts. Thus, corporate law has evolved towards offering managers a blanket safe harbor that applies within very broad conditions.

A final path the law can take in the face of biased ex post adjudicators is to create causes of action that need to be litigated prior to the materialization of the outcome.

Under such regimes courts engage in a substantive evaluation of the decision, much like in any negligence case. However, this evaluation is conducted *ex ante*, before harm had a chance to materialize.

The Israeli regime regulating the payment of dividends can serve as a case in point. The payment of dividends by the corporation is a means through which the assets of the corporation are transferred to the hands of its shareholders. In this way, dividends create a conflict of interests between shareholders and creditors. Whereas the former have a strong interest in transferring wealth to their pockets, the latter have an interest in protecting the corporation's liquidity so it can continue to serve its debt. Given this conflict of interests, corporate law often regulates the distribution of dividends, and limits the ability of shareholders to extract cash from the company. Under Israeli law a corporation must fulfill two criteria in order to move forward with a dividend (§302 of the Corporate Code): (1) a rigid rule-like profit test that examines whether from an accounting perspective the corporation has profits on its balance sheet; and (2) a flexible standard-like solvency test that examines whether the payment of the dividend will create an unreasonable risk that the company will not be able to fulfill its obligations.

Courts evaluating the solvency test in cases in which the corporation became insolvent might find it difficult to ignore this fact. As a result, they are expected to systematically view legitimate dividend payments as payments that did not fulfill the solvency test. Fearing the personal liability associated with a finding that a corporation distributed a dividend in violation of Section 302, directors are expected to adopt defensive policies and approve an inefficiently low level of dividends.

A partial solution to this concern is provided by §303 of the Corporate Code. According to this section in some cases a corporation can approach the court and ask it to rule in advance that it fulfills the solvency test and can go forward with the payment of a dividend. This provision enables the corporation to force its creditors to litigate over the murky solvency question *ex ante*. By doing so it can sidestep any biased *ex post* disputes over the matter. As was the case with respect to negligence and regulation, however, adopting such a regime might generate significant enforcement costs.

## 5.5 Contracting out of the Hindsight Bias

Thus far the analysis has focused on the way the law can deal with the problems created by the hindsight bias. Contracting parties who anticipate the influence of the bias on litigation that might arise down the road do not have to rely on legislated solutions. Rather, they can design their contract such that it will minimize the costs generated by the bias. By adopting clear *ex ante* norms in the contract the parties reduce the possibility of biased judgments *ex post*.

An example of an attempt to deal with the bias through contract design can be found in the context of remedies for breach. The default remedy in Anglo-American contract law is expectation damages (Farnsworth 2004, §112.1, at 149; Peel and Treitel 2007, §20-054, at 1024). Generally speaking, this remedy is limited to losses that the

breaching party could foresee at the time of contracting as a probable result of the breach (*Hadley v. Baxendale* 1854). Thus, much like in other areas of law, courts determining contractual remedies need to assess the remedial question based on the information set that was available at the time of contracting. Given the existence of the hindsight bias, courts are expected to systematically overestimate the foreseeability of losses.

Contracting parties who anticipate the skewed pattern of judicial decisions relating to contract remedies, however, do not have to take it as a given. Rather, they can attempt to avoid them by opting out of the remedial regime offered by the law. More specifically, they can stipulate the level of damages that will be applied in case of breach in the contract. By doing so the parties limit the power of biased adjudicators, and increase the likelihood that in the case of a breach the court will apply the remedy that best fits their *ex ante* expectations.<sup>2</sup>

To be sure, stipulating damages *ex ante* is not a foolproof solution to the problems associated with hindsight. Even when the parties agree to a specific sum of damages, the court can still strike down this sum if it finds that it is not “an amount that is reasonable in the light of the anticipated or actual loss caused by the breach and the difficulties of proof of loss” (*Restatement (Second) of Contracts*, §356). While this evaluation might be somewhat constrained by the predetermined sum (that could even function as an anchor), there is no reason to assume that it will be immune from the bias. Apparently, you can run but you cannot hide from the hindsight bias.

## 6 CONCLUSION

---

This chapter examined the way in which the hindsight bias influences the decisions made by judges and jurors. The general picture arising out of it is that the bias affects a wide body of legal decisions ranging from tort law to constitutional rights. While the positive predictions associated with the bias are clear, its normative implications are a much more murky business. As the analysis presented in the chapter suggests, the fact that courts are reaching biased decisions *ex post* does not necessarily imply that those decisions are inefficient. A complete welfare analysis of the bias requires accounting for both the inefficiencies of the legal rule itself, and other biases that come into play when it is applied. In this regard, the chapter has mapped numerous paths for future research.

<sup>2</sup> The text focuses on the role of the hindsight bias on opting out of the default remedial regime, though there are clearly other factors that might come into play such as the parties' attitude towards risk, their reliance decisions, etc.

## ACKNOWLEDGMENTS

---

I thank Eyal Zamir and participants in the conference on behavioral law and economics at Notre Dame Law School for their helpful comments, and Nathan Johnson for his valuable research assistance. Financial support for this study was provided by the Israel Science Foundation (grant 1283/11) and by the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation (grant No. 1821/12).

## REFERENCES

---

- Anderson, J. C., Jennings, M. M., Kaplan, S. E., and Reckers, P. M. 1995. The effect of using diagnostic decision aids for analytical procedures on judges' liability judgments. *Journal of Accounting and Public Policy*, 14: 33–62.
- Baron, J., and Hershey, J. C. 1988. Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, 54: 569–79.
- Bernstein, M. H. 1955. *Regulating Business by Independent Commission*. Princeton, NJ: Princeton University Press.
- Blumstein, J. F. 2006. Medical malpractice standard-setting: Developing malpractice safe harbors as a new role for QIOs. *Vanderbilt Law Review*, 59: 1017–50.
- Burk, D. L., and Lemley, M. A. 2009. *The Patent Crisis and How the Courts Can Solve It*. Chicago: University of Chicago Press.
- Caplan, R. A., Posner, K. L., and Cheney, F. W. 1991. Effect of outcome on physician judgments of appropriateness of care. *JAMA: The Journal of the American Medical Association*, 265: 1957–60.
- Casper, J. D., Benedict, K., and Kelly, J. R. 1988. Cognitions, attitudes, and decision-making in search and seizure cases. *Journal of Applied Social Psychology*, 18, 93–113.
- Casper, J. D., Benedict, K., & Perry, J. L. 1989. Juror decision making, attitudes, and the hindsight bias. *Law and Human Behavior*, 13: 291–310.
- Cheney, F. W., Posner, K., Caplan, R. A., and Ward, R. J. 1989. Standard of care and anesthesia liability. *JAMA: The Journal of the American Medical Association*, 261: 1599–603.
- Christensen-Szalanski, J. J., and Willham, C. F. 1991. The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 48: 147–68.
- Dawson, N. V., Arkes, H. R., Siciliano, C., Blinkhorn, R., Lakshmanan, M., and Petrelli, M. 1988. Hindsight bias: An impediment to accurate probability estimation in clinicopathologic conferences. *Medical Decision Making*, 8: 259–64.
- Farnsworth, E. A. 2004. *Farnsworth on Contracts*. 3rd ed. Vol. 3. New York: Aspen Publishers.
- Feldman, Y., and Teichman, D. 2009. Are all legal probabilities created equal? *New York University Law Review*, 84: 980–1022.
- Fischhoff, B. 1975. Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1: 288–99.
- Fischhoff, B. 1977. Perceived informativeness of facts. *Journal of Experimental Psychology, Human Perception and Performance*, 3: 349–58.

- Gino, F., Shu, L. L., and Bazerman, M. H. 2010. Nameless + harmless = blameless: When seemingly irrelevant factors influence judgment of (un)ethical behavior. *Organizational Behavior and Human Decision Processes*, 111: 93–101.
- Graham v. John Deere Co.* 1966. 383 U.S. 1.
- Guilbault, R. L., Bryant, F. B., Brockway, J. H., and Posavac, E. J. 2004. A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology*, 26: 103–17.
- Guthrie, C., Rachlinski, J. J., and Wistrich, A. J. 2000. Inside the judicial mind. *Cornell Law Review*, 86: 777–830.
- Guthrie, C., Rachlinski, J. J., and Wistrich, A. J. 2007. Blinking on the bench: How judges decide cases. *Cornell Law Review*, 93: 1–44.
- Hadley v. Baxendale.* 1854. 9 Exch. 341, 156 Eng. Rep. 145.
- Hastie, R., and Viscusi, W. K. 1998. What juries can't do well: The jury's performance as a risk manager. *Arizona Law Review*, 40: 901–21.
- Hawkins, S. A., and Hastie, R. 1990. Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, 107: 311–27.
- Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., and Muller, M. 1988. Hindsight bias: An interaction of automatic and motivational factors? *Memory and Cognition*, 16: 533–38.
- Hertwig, R., and Erev, I. 2009. The description-experience gap in risky choice. *Trends in Cognitive Sciences*, 13: 517–23.
- Jolls, C. 1998. Behavioral economics analysis of redistributive legal rules. *Vanderbilt Law Review*, 51:1653–1677.
- Jolls, C., Sunstein, C. R., and Thaler, R. 1998. A behavioral approach to law and economics. *Stanford Law Review*, 50: 1471–550.
- Kamin, K. A., and Rachlinski, J. J. 1995. Ex post ≠ ex ante: Determining liability in hindsight. *Law and Human Behavior*, 19: 89–104.
- Klass, A. B. 2009. Tort experiments in the laboratories of democracy. *William and Mary Law Review*, 50: 1501–76.
- Korobkin, R. B., and Ulen, T. S. 2000. Law and behavioral science: Removing the rationality assumption from law and economics. *California Law Review*, 88: 1051–144.
- LaBine, S. J., and LaBine, G. 1996. Determinations of negligence and the hindsight bias. *Law and Human Behavior*, 20: 501–16.
- Loftus, E. F., and Beach, L. R. 1982. Human inference and judgment: Is the glass half empty or half full. *Stanford Law Review*, 34: 939–56.
- Lord, C. G., Lepper, M. R., and Preston, E. 1984. Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47: 1231–43.
- Lowe, D. J., and Reckers, P. M. 1994. The effects of hindsight bias on jurors' evaluations of auditor decisions. *Decision Sciences*, 25: 401–26.
- Mandel, G. 2006. Patently non-obvious: Empirical demonstration that the hindsight bias renders patent decisions irrational. *Ohio State Law Journal*, 67: 1391–463.
- Mandel, G. 2006–7. Patently non-obvious II: Experimental study on the hindsight issue before the Supreme Court in *KSR v. Teleflex*. *Yale Journal of Law and Technology*, 9: 1–43.
- Nario, M. R., and Branscombe, N. R. 1995. Comparison processes in hindsight and causal attribution. *Personality and Social Psychology Bulletin*, 21: 1244–55.
- Parchomovsky, G., and Stein, A. 2008. Torts and innovation. *Michigan Law Review*, 107: 285–315.
- Peel, E., and Treitel, G. H. 2007. *The Law of Contract*. 12th ed. London: Sweet and Maxwell.



- Poythress, N. G. 1989. Negligent release litigation: A proposal for procedural reform. *Journal of Psychiatry and Law*, 17: 595–606.
- Rachlinski, J. J. 1998. A positive psychological theory of judging in hindsight. *University of Chicago Law Review*, 65: 571–625.
- Rachlinski, J. J., Guthrie, C., and Wistrich, A. J. 2011. Probable cause, probability, and hindsight. *Journal of Empirical Legal Studies*, 8: 72–98.
- Roese, N. J., and Vohs, K. D. 2012. Hindsight bias. *Perspectives on Psychological Science*, 7: 411–26.
- Sanna, L. J., and Schwarz, N. 2003. Debiasing the hindsight bias: The role of accessibility experiences and (mis)attributions. *Journal of Experimental Social Psychology*, 39: 287–95.
- Sanna, L. J., Schwarz, N., and Stocker, S. L. 2002. When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28: 497–502.
- Shavell, S. 1985. Uncertainty over causation and the determination of civil liability. *Journal of Law and Economics*, 28: 587–609.
- Shavell, S. 2012. A fundamental enforcement cost advantage of the negligence rule over regulation. National Bureau of Economic Research Working Paper no. w18418.
- Stallard, M. J., and Worthington, D. L. 1998. Reducing the hindsight bias utilizing attorney closing arguments. *Law and Human Behavior*, 22: 671–63.
- Stein, A., and Parchomovsky, G. 2010. The distortory effect of evidence on primary behavior. *Harvard Law Review*, 124: 518–48.
- Taragin, M. I., Willett, L. R., Wilczek, A. P., Trout, R., and Carson, J. L. 1992. The influence of standard of care and severity of injury on the resolution of medical malpractice claims. *Annals of Internal Medicine*, 117: 780–84.
- Tetlock, P. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?*. Princeton, NJ: Princeton University Press.
- Turner, M. A. 2009. Recent developments in Indiana tort law. *Indiana Law Review*, 43: 1053–88.
- Vars, F. E. 2010. Toward a general theory of standards of proof. *Catholic University Law Review*, 60: 1–50.
- Viscusi, W. K. 1999. How do judges think about risk? *American Law and Economics Review*, 1: 26–62.
- The Walt Disney Co. Derivative Litigation*. 2006. 906 Del. A.2d 27.
- Wistrich, A. J., Guthrie, C., and Rachlinski, J. J. 2005. Can judges ignore inadmissible information? The difficulty of deliberately disregarding. *University of Pennsylvania Law Review*, 153: 1251–345.
- Zamir, E., and Ritov, I. 2012. Loss aversion, omission bias, and the burden of proof in civil litigation. *Journal of Legal Studies*, 41: 165–207.



PART IV

---

**BEHAVIORAL  
ECONOMICS**  
*Legal Applications*

---



## CHAPTER 15

---

# BEHAVIORAL LAW AND ECONOMICS OF PROPERTY LAW

## *Achievements and Challenges*

---

DAPHNA LEWINSOHN-ZAMIR

### 1 INTRODUCTION

---

ECONOMIC analysis of property law occupies a central place in the property literature. One might even say that standard economic analysis has reached its zenith and to a certain extent exhausted its usefulness in addressing real-life property issues. After laying down the main economic justifications for private property (Posner 2011, pp. 40–42), and deriving basic rules—such as those relating to takings compensation—from these ideas (e.g., Michelman 1967), scholars have turned to ever more sophisticated and complex analyses that are less applicable to actual property conflicts (see Lewinsohn-Zamir 2001b, pp. 98–101). In contrast, behavioral analysis of property law is still relatively novel. Research in this field is in the early stage of testing or applying basic insights and theories, and currently covers rather limited or sporadic issues.

THIS chapter critically surveys the behavioral law and economics literature on property law. Section 2 briefly sketches the main features and tendencies of existing studies. Section 3 discusses several issues with respect to which behavioral analysis has made an impact, such as the characterization of property as a “thing” or as a “bundle of sticks,” compensation for takings of property, the choice between property rules and liability rules, and redistribution. In addition, section 3 highlights gaps in the current literature, such as the disregard of commercial property, fungible property, and movables. Section 4 concludes the survey with remarks on the prospects of debiasing, and recommendations for future research.

## 2 BEHAVIORAL ANALYSIS OF PROPERTY LAW: AN OVERVIEW

---

The behavioral literature on property law can be divided into two categories that partially correspond with two different periods. One category consists of (mostly earlier) studies that did not conduct independent experimental research, but rather relied on general results from the vast psychological literature. Scholars applied findings and theories that were developed in nonlegal contexts to the analysis of law-related issues. These applications were based on common-sense reasoning and analogies, rather than on direct testing of a property law topic. Thus, Fennell (2003) used phenomena like the omission and regret-avoidance biases (see Ritov and Baron 1990, 1992; Baron and Ritov 1994), to explain why, despite the popular opposition to estate taxes, people do not dispose of their property during their lifetime. Similarly, Lewinsohn-Zamir (2008, pp. 661–65, 681–87) relied on these biases to justify the limited legal intervention in owners' freedom to destroy their property. In a similar vein, Stern (2006) employed the crowding-out effect (see Gneezy and Rustichini 2000) to support certain financial incentives that would safeguard landowners' intrinsic motivation to conserve wildlife habitats, thereby increasing the efficacy of conservation programs. Another example is Tor and Oliar's (2002) reliance on evidence regarding people's overoptimism (see Weinstein 1980, 1987) to argue that a copyright protection term phrased as "lifetime of the author plus X years after her death" would be perceived as granting larger financial incentives to create, than a fixed term of years of comparable expected duration.

The second category or phase of behavioral studies of property law is characterized by endeavors to carry out experimental research explicitly tailored to legal issues. Scholars have either designed their own experiments or relied on studies that directly examined law-related topics. This second wave of research responded to calls in the literature to improve the relevance or "fit" of the data employed in behavioral analyses (Barros 2009, pp. 646–47). Accordingly, Nadler and Diamond (2008) investigated the adequacy of market-value compensation for the expropriation of residences, by expressly asking subjects about the sum of money above market value that would be required for them to sell their home (with the understanding that if negotiations fail, the government would exercise its eminent domain powers to compel the transfer). In a similar fashion, Lewinsohn-Zamir (2013) directly elicited laypersons' and businesspeople's preferences regarding monetary and in-kind remedies for various types of injury, including trespass to land and partial expropriation of a parcel.

Along with their important contribution to property law scholarship, both types of behavioral studies suffer from considerable limitations that may be due to the nascent stage of this theoretical and methodological perspective. As behavioral analyses of property law grow in quantity and sophistication, the limitations described below will hopefully disappear, or at least be reduced.

Behavioral studies tend to test the existing rules of property law. In itself, this is certainly a worthwhile enterprise. However, excessive focus on debates regarding the prevailing rules has its disadvantages as well. For instance, a substantial part of the literature engages with the quintessential nonfungible property—the home—and whether it indeed merits special legal protection (Barros 2006; Blumenthal 2009, pp. 617–21, 640–41; Godsil 2004; Kelly 2006; Nadler and Diamond 2008; Stern 2009, 2010). Behavioral analyses have largely overlooked nonresidential lands (such as commercial units) and other types of property (such as movables). Consequently, these studies have not examined the extent to which certain psychological findings are relevant to fungible property as well.

Another difficulty is that legal studies tend to rely on general behavioral phenomena, without paying sufficient attention to context-specific variations, exceptions, and refinements established in empirical and experimental psychological studies. Thus, for example, the basic endowment effect is widely applied in the literature (see, e.g., Ayres 1998; Buccafusco and Sprigman 2010, 2011; Jolls and Sunstein 2006), but the circumstances in which it is eliminated or reduced are often overlooked. These circumstances may be relevant to the legal issue at stake.

The final characteristic to be noted here is that legal research commonly strives to be normative as well as descriptive. It does not suffice with establishing experimental results, but tends to deduce normative implications from the data. These policy suggestions are sometimes problematic. Typically, the experimental or empirical results are insufficient, in themselves, to support any specific legal recommendation; and sometimes they may even support conflicting recommendations. Quite often, the psychological findings can be regarded, at best, as one consideration among many. Another obstacle is the level of generality of behavioral data. The data may be inapplicable to legal issues because they do not test the nuances that are important for legal analysis. For instance, while the distinction between ownership and possession is crucial in various legal contexts—ranging from adverse possession, to self-help and landlord-tenant relationships—psychological studies typically fail to distinguish between ownership and possession. Finally—as is often the case when trying to infer from laboratory findings to the real world—drawing conclusions from experimental evidence and applying them to legal policymaking may be problematic due to the ineliminable differences between “property” in the laboratory and in real-life scenarios.

The next section of the chapter demonstrates the above tendencies in an array of property contexts.



### 3 BEHAVIORAL ANALYSIS OF PROPERTY ISSUES

---

Behavioral studies have made significant contributions to several debates in property law. In what follows, I survey some of the main issues on which the psychological findings have left their mark. These include the characterization of property as a “thing” or as a “bundle of sticks,” the distinction between fungible and nonfungible property and between ownership and possession, the choice between property rules and liability rules, and the use of substantive rules of property law for redistributing welfare.

#### 3.1 Property as Thing versus Property as Bundle of Sticks

Property scholars debate whether property should be regarded as a “thing” or as a “bundle of sticks” (Dagan 2011, ch. 2; Merrill and Smith 2007; Penner 1996; Smith 2012). The former metaphor views property as a discrete object, whereas the latter portrays it as a package of separate rights: to possess, use, exclude others, dispose of, destroy, and so forth (for a summary of the literature on the two property paradigms, see Nash 2009, pp. 694–707). This controversy is not a matter of semantics. Rather, it is believed that the framing of property one way or the other affects people’s perceptions of ownership, and consequently the legal protection that the state affords owners.

Specifically, some writers assert that the “thing” conception implies exclusive, unlimited control over assets. Accordingly, any restriction of this right is an injury to property, and such a perception might eventually lead to excessive protection of owners at the expense of other social goals. A bundle of rights understanding, in contrast, draws one’s attention to the limitations of ownership and to the interests of others. Consequently, state interference with property would be more readily accepted as legitimate (Nash and Stern 2010, pp. 451–55, 462–65; Singer 2000, pp. 2–13).

As a theoretical generalization, these arguments seem unpersuasive. In principle, why can’t people regard property as a thing conferring limited rights, or as a bundle of unrestricted rights? Furthermore, one may conjecture that the two property perceptions would have an opposite effect. Indeed, in theoretical discussions of the takings issue, some scholars have argued that if property is perceived as a thing, fewer state interventions would be considered to be takings. This is because under the “thing” conception, property is likely to be regarded as taken only when it is physically expropriated or when nonphysical regulation renders it useless. That is to say, as long as some reasonable use of the asset remains, the owner would not think that her “thing” was taken from her. In contrast, the bundle of sticks conception emphasizes each distinct right in the ownership-package. As a result, eliminating any one of these sticks would be viewed as a taking of that right in its entirety. This argument was first advanced by Ackerman (1977, ch. 6), who claimed that laypeople (“ordinary observers”) adhere to the thing conception of property whereas legal professionals (“scientific

policymakers”) view property as a bundle of rights. Similar views regarding the probable effect of a “bundle of sticks” understanding were expressed by property theorists like Michelman (1988, pp. 1614–25) and Radin (1988, pp. 1676–78).

Interestingly, behavioral scholars attempted to test the former effect of property perception but did not mention the possibility of the latter, opposite effect. Nash (2009) examined how the framing of a laptop computer as a thing or as a bundle of rights affected in-coming law students’ acceptance of restrictions on the asset. Subjects were told to envision a scenario in which they had to purchase a laptop from the law school. Half of the questionnaires tried to frame the laptop as a thing (e.g., by referring to the acquisition of “a laptop”) and the other half as a bundle of rights (e.g., by indicating that the subjects purchase the right to use a laptop). In both cases, subjects stated their attitudes toward the law school’s proposed limitations on the use of the laptops (for instance, with respect to the hours and types of permitted uses). Nash found that while all subjects expressed unfavorable reactions to the suggested restrictions, students who received the bundle version of the questionnaire were more accepting of the law school’s intervention than students who answered the thing version (Nash 2009, pp. 715–21).

The findings that the notion of property can be manipulated, and that a bundle framing weakens ownership perceptions and decreases resistance to regulation of the property, was replicated and further advanced by Nash and Stern (2010), who similarly used vignettes regarding students’ rights in laptops purchased from their law school. Both studies advocate reframing people’s property perceptions from “thing” to “bundle” in various legal contexts, such as environmental regulation (e.g., Nash and Stern 2010, pp. 492–94). In support of this recommendation, Nash and Stern (2010, p. 481) claim that the public acceptance of zoning regulation—in contrast to the popular hostility towards eminent domain—is an example of successful bundle-of-rights reframing. However, one may conversely argue that if people indeed view zoning as legitimate, then this phenomenon is actually more supportive of the opposing theory: since people perceive their property as a *thing*, they do not regard it as taken by the state if it remains physically in their hands and can still be used in economically viable ways.

While these experimental studies are important, they do not settle the debate regarding the effect of different property perceptions, and their normative implications may be challenged. First, one may argue that the bundle-of-sticks questionnaires did not invoke in subjects a notion of “having property” at all, because they expressly gave them only a set of use rights (see also Nash 2009, pp. 721–22). This possibility is reinforced by the fact that the subjects did not receive actual laptops, but only answered hypothetical questions regarding the asset. Second, since the laptops were purchased from the law school (not from a third party), the subjects may have viewed the restrictions as a “limited giving” of property rather than as a “taking,” and consequently regarded the interference as more legitimate than otherwise. Third, even if property conceptions can be successfully manipulated in experimental settings, this might not be possible in real life. One reason for skepticism is that people may not be aware in advance of restrictions on their property rights, or may not actually experience them

with respect to most of their property. Therefore, it is doubtful if we can educate them to perceive property as a bundle of specific and limited rights, rather than as discrete assets. Finally, and more importantly, it is difficult to deduce normative implications from people's perceptions of property, whatever they may be. It may be the case that property conceptions play a rather small role in owners' reactions to restrictions on their assets, and that the context and substance of these limitations are more crucial. Thus, the type of asset involved, the magnitude of the injury to the asset's value, and the extent of the distribution of the injury throughout society (whether the injury is inflicted on a small group of property owners or widely distributed among many), may be better indicators of the probable impact of property restrictions than the property perception of the injured owners (Lewinsohn-Zamir 1996).

### 3.2 Fungible and Nonfungible Property

Property can be classified as fungible or nonfungible. This distinction figures prominently in Margaret Radin's influential personhood theory of property (Radin 1982). Radin based her theory on Hegel's justification of private property as necessary for the development of people's personality. According to the personhood theory, property is "personal" if its loss cannot be remedied by receiving its value or purchasing a replacement in the market (paradigmatic examples are the home, a wedding ring, or a family portrait). A fungible asset, in contrast, is easily replaceable with a similar object (e.g., money, a contractor's parcel of undeveloped land, or a commercial landlord's apartment). Generally speaking, Radin has argued that greater protection should be afforded to personal property than to fungible property. For instance, preventing even the compensated taking of a highly personal asset, such as the family home, may be justified. In contrast, monetary compensation should always suffice for claims involving fungible property, and sometimes no compensation is necessary (Radin 1982, pp. 959–61, 988, 1005–6, 1014–15; 1986, pp. 362–65; 1988, pp. 1868–92).

Over the years, nonbehavioral studies have criticized these normative conclusions. Scholars have argued that property deemed personal according to personhood theory is granted excessive legal protection. Such protection has elitist and regressive consequences because it favors people with large quantities of expensive personal property, like luxury homes (Schnably 1993, pp. 375–79, 397–99). In certain contexts, such as bankruptcy proceedings, it suffices to guarantee individuals some reasonable shelter, rather than allowing them to remain in their actual place of residence, regardless of its value (Lewinsohn-Zamir 2003, pp. 1721, 1725–29). It was also averred that fungible property receives inadequate protection under personhood theory. The fact that money can sufficiently compensate for injury to fungible property does not imply that monetary compensation is unnecessary (Lewinsohn-Zamir 1996, p. 121).

The behavioral legal literature has focused mainly on the "personal" strand of this debate and, particularly, on the protection of homes. It inquires whether homes are indeed "special" (Barros 2006, 2009), and overlooks the controversy regarding the

adequate protection of fungible property, such as land held for business purposes. Furthermore, property other than land—whether personal or nonpersonal—is mostly ignored.

Nadler and Diamond (2008), for example, conducted an experimental study on the expropriation of homes. Subjects were asked to indicate the sum of money above fair market value that would be required for them to sell their home, with the understanding that if negotiations for a voluntary sale were to fail, the government would compel the transfer. The authors found that respondents indeed required compensation surpassing market value and that the length of time the residence had been in the family (two years versus one hundred years) significantly affected the difference between market value and the hypothetical sale price (Nadler and Diamond 2008, pp. 729–30, 743–44).<sup>1</sup> However, they did not examine whether similar phenomena exist with respect to other types of property, such as commercial land.

Focusing on the same issue, Stern (2009) relied on psychological research to critique the presumed uniqueness of homes. Some studies have shown that psychological attachment to the home is weaker than commonly believed. For instance, most people forced to relocate from their homes due to urban-renewal projects eventually acclimated to their new surroundings and did not suffer serious long-term mental harm. Therefore, Stern advocated reducing the number and scope of home-protective laws (Stern 2009, pp. 1115–19, 1139–44). Likewise, Stern criticized the expansive constitutional protection of residences from search and seizure. Psychological studies indicate that privacy is not a rigid spatial concept and empirical evidence shows that homes are commodities that are sold or rented quite frequently. Hence, Stern argued that the strong emphasis on the physical home should be replaced with rules that are more responsive to the concerns of substantive privacy (Stern 2010, pp. 906–11, 923–30).

The focus on homes is understandable. A person's residence is often her most valuable asset, and one that is relatively vulnerable to injury by the state. While a home might be expropriated for public use (e.g., *Kelo v. New London*, 125 S.Ct. 2655 (2005)), it is difficult to envision a scenario where a wedding ring or a family portrait would be taken by the government. However, the disproportionate attention to homes leaves important questions unanswered. Can people form strong attachments to property regarded as “fungible”? Does monetary compensation or fair market value always suffice when fungible property is injured? More specifically, should commercial property necessarily receive less protection than residential?

Intuitively, it is not clear why property used for commercial purposes should be classified as “nonpersonal.” Many people spend most of their time and efforts in their business, and people's occupation is often central to their self-identity. Plausibly, individuals' welfare depends on their business's flourishing, through which they realize their talents, independence, and personal character. Moreover, even if owners indeed

<sup>1</sup> In contrast, the public purpose of the taking—building a children's hospital or rather a shopping mall—had little effect on hypothetical sale prices. Nadler and Diamond 2008, pp. 742–43.

regard commercial assets as fungible, jurists should be wary of drawing conclusions from this fact in the absence of sufficient behavioral data about fungible property. Let me illustrate this concern in the context of compensation for takings of property.

In the United States (and elsewhere), there is a sharp distinction between physical and nonphysical injuries to land. Landowners are afforded much wider protection against physical injuries (such as the taking of possession or the elimination of existing uses on the land) than against nonphysical injuries (such as the restriction of unrealized development rights or the prohibition of future uses). Whereas in the former case even a small injury requires compensation, in the latter case enormous reductions in value are legitimized without compensation (Lewinsohn-Zamir 1996, pp. 114–19).<sup>2</sup> Radin (1982, pp. 1007–8; 1988, pp. 1691–92) explained this rule of noncompensation for severe value reductions in that development rights are fungible property, and that the owner often holds the injured land solely for investment or commercial purposes. Other scholars strengthened this justification by relying on the psychological literature on the endowment effect (EE) (see generally chapter 12 by Korobkin in this volume).

Numerous experiments have shown that people value an entitlement they already possess much more than an identical entitlement that they have an opportunity to acquire (Camerer 1995, pp. 665–70; Kahneman, Knetsch, and Thaler 1990, 2008; Knetsch and Sinden 1984). A major explanation for the EE is loss aversion. Parting with an entitlement is perceived as a loss, whereas acquiring the same entitlement is viewed as a gain. Since losses loom larger than gains, people's selling price is significantly higher than their purchase price (Camerer 1995, p. 668; Kahneman 1992; Korobkin 2003, pp. 1250–55; Zamir 2012, pp. 835–40; chapter 11 by Zamir in this volume).<sup>3</sup>

Based on this phenomenon, Ellickson rationalized that a physical injury to land would ordinarily be viewed as a loss. In contrast, a nonphysical injury is likely to be perceived as unattained gains. Therefore, the former type of injury is more worthy of protection than the latter (Ellickson 1989, pp. 35–38). In a similar vein, Serkin claimed that elimination of an existing use is experienced differently than the prohibition of a future use—the former as an out-of-pocket cost and the latter as a forgone gain. Consequently, existing uses deserve more protection than prospective uses like development rights (Serkin 2009, pp. 1267–69). Both explanations justify the physical/nonphysical distinction by linking restriction of development rights with people's psychological reaction to unattained profits. These arguments are based on a crucial assumption, namely, that nonphysical injuries are indeed perceived as forgone gains. However, the opposite may well be true. Why wouldn't landowners experience downzoning as a loss of formerly vested development rights? This presumption seems especially plausible if the land's

<sup>2</sup> The physical/nonphysical distinction is not a universal phenomenon. Countries such as Germany and Israel recognize a broad right to compensation for both types of injury. For a comparative study of land use regulation and compensation in several countries, see Alterman 2010.

<sup>3</sup> But see Plott and Zeiler (2005, 2007), who argue that the EE is due to subject misconceptions caused by the specific procedures used to elicit valuations. For criticism of these studies, see Isoni, Loomes, and Sugden 2011; Korobkin's chapter in this volume.

purchase price reflected the value of the development rights. Absent direct testing of the matter, one cannot presume one way or the other. Moreover, even if restriction of development rights is viewed as an unattained gain—and thus the injury to the landowner is smaller than if such restriction were regarded as a loss—the injury may still be substantial enough to warrant compensation.

In contrast to the arguments of Ellickson and of Serkin, Fischel acknowledges that development rights can create an EE and that their limitation may be perceived as a loss. However, Fischel criticizes the claim that the EE supports takings compensation exceeding fair market value. This claim, so the argument goes, ignores the fact that compensation payments are financed by taxpayers and that the latter too may experience an EE when parted from their income (Fischel 1995, pp. 207–10). Fischel's critique is problematic because, in the absence of experimental evidence, one cannot assume that the two effects offset each other. The EE from loss of development rights may be substantially larger than the EE from loss of money (especially when taxes are fully or partially hidden, as in the case of employers deducting social security payments from salaries [McCaffery 1994, pp. 1874–86]). A possible reason for such disparity is that the loss of development rights relates to a specific asset, whereas a monetary loss from taxes affects a person's wealth in general. In addition, while the former type of injury typically affects only a certain group of landowners, the latter type is widespread across all taxpayers. If this is indeed the case, overmarket compensation for takings of property could still be justified.

Notwithstanding the scholarly focus on nonfungible property and, particularly, on homes, there is some evidence that the fungible/nonfungible dichotomy is overly rigid and might lead one astray. Psychological studies indicate that people form attachments to fungible property. Beggan (1992) discovered a “mere ownership” effect. Individuals are biased in favor of an object and rate it as more attractive just because they own it (and regardless of whether it is about to be transferred). Beggan argued that people desire to maintain a positive self-image and therefore overvalue the objects they own (Beggan 1992, p. 235). This desire to enhance one's positive image arises, in particular, when the said image is threatened (such as by subjects' failure in an assigned task). That is to say, a decrement in self-image increases the magnitude of the mere ownership effect (Beggan 1992, pp. 233–34). Importantly for our purposes, the ownership effect was found with respect to cheap fungible assets, such as a cold-drink insulator.<sup>4</sup>

In a similar vein, Belk (1988) offered extensive evidence to support the claim that people generally regard their possessions as part of their identity or extended self. The perception of possessions as part of one's self is not limited to objectively unique assets but extends to fungibles, including money. This phenomenon is created in various ways, such as through controlling, creating, or knowing an asset (Belk 1988, pp. 150–51, 155).

<sup>4</sup> The “mere ownership effect” is supported by experiments in which the purchase price of buyers who happened to own an object identical to the one they were offered to buy equaled seller's asking price. See Morewedge et al. 2009.



Thus, a regular bicycle may invoke pleasant memories, a standard chair may be associated with a loved person who sat in it, and ordinary athletic equipment may symbolize skills and talents. Belk (1988, pp. 142–43) explains that involuntary loss of such possessions—for example, by theft, natural disaster, or forced disposition due to economic necessity—causes trauma and loss of self; voluntary disposition, in contrast, does not produce this effect. Csikszentmihalyi and Rochberg-Halton (1981) also established the connection between mundane assets and the development of the self. As they explain, objects affect what a person can do, and since what a person does is largely what he or she is, objects have a determining effect on self-development (Csikszentmihalyi and Rochberg-Halton 1981, p. 53). Thus, for instance, “the tools of one’s trade. . . help to define who we are as individuals” (Csikszentmihalyi and Rochberg-Halton 1981, p. 92).<sup>5</sup>

The “fungibility” of seemingly fungible assets was also questioned in legal contexts. Lewinsohn-Zamir (2013) tested the common belief that people are indifferent between in-kind and monetary remedies of equal pecuniary value. It was found that both laypersons and experienced businesspeople strongly prefer in-kind entitlements and remedies over monetary ones. Moreover, the more experienced the businesspeople were, the greater their reluctance to content themselves with a monetary remedy. Importantly, a preference for in-kind entitlements and remedies existed even when the right-holder was a firm, and even when the remedy was related to fungible, easily replaceable assets, whose market value was ascertainable. For instance, in contradiction to Radin’s theory, subjects did not view development rights as fungible property: in a vignette depicting the expropriation of 10 percent of an undeveloped parcel to widen an adjacent road and sidewalk, they clearly preferred compensation in development rights to the remainder of the parcel over money compensation of equal and even substantially higher value (Lewinsohn-Zamir 2013, pp. 163–65, 171–74, 186–88). Another experimental study (Lewinsohn-Zamir 2012) found that both laypersons and businesspeople perceive outcomes broadly and therefore do not confine the judgment of events to their end results. Certain factors, such as how an outcome was brought about (e.g., in a spirit of goodwill and mutual cooperation or not), the identity of the parties involved (e.g., strangers or friends), the voluntariness or nonvoluntariness of their behavior, and the intentionality or nonintentionality of their acts, are commonly regarded as part of the outcome and significantly affect its valuation. Thus, for example, the very fact that an asset was transferred unwillingly inflicts a loss that is distinct from the owner’s subjective valuation of the asset. This was demonstrated even in the case of a vacant parcel expropriated from a real-estate company that held it as part of its commercial stock and valued it at the current market price (Lewinsohn-Zamir 2012, pp. 872, 876, 882–83). One implication of these studies is that the commonly awarded “fair market value” compensation may systematically undercompensate property owners even when the

<sup>5</sup> For a study of self-extension and self-expression through workplace possessions, see Tian and Belk 2005.



injured asset is fungible (Lewinsohn-Zamir 2013, pp. 178–88; Lewinsohn-Zamir 2012, pp. 888–90).

To sum up, behavioral studies cast serious doubt on the prevailing distinction between fungible and nonfungible property. This dichotomy is overly rigid since potentially, any asset could be viewed as personal (or nonpersonal) by its owner. Furthermore, even when an asset is subjectively perceived as fungible, market-value damages may fail to compensate for the injury resulting from its forced taking. To be sure, the legal system may have good reasons not to adopt people’s perceptions in this regard. For instance, the unwillingness to encourage fetishism or the high costs of tailoring compensation rules may sometimes limit redress to fair market value. However, if one holds that enhancing human welfare is a central goal of legal policymaking, and if one espouses a theory of welfare that is concerned with preference fulfillment, then it is important to be aware of the flaws of the fungible/nonfungible distinction, and particularly its underestimation of “ordinary” property and assets used for commercial purposes.<sup>6</sup>

### 3.3 Ownership and Possession

Ownership and possession are basic building blocks of property law, and every so often they vest in the same person. Legal scholars, however, are well aware of the importance of the distinction between ownership and possession, and of the fact that the two may diverge. For example, landlords are nonpossessing owners of the tenement, whereas tenants are nonowning possessors of it. In addition, jurists acknowledge the significance of distinguishing between lawful and unlawful possession, as well as between actual and prospective possession. Thus, a case of adverse possession may involve a nonpossessing owner and an unlawful possessor; and a trespass scenario may pit a person with a right to receive possession against a person who possesses in fact. Indeed, conflicts between owners and possessors—both lawful and unlawful—are not uncommon.

The legal treatment of such situations could benefit from behavioral research. Moreover, the relevance of psychological data goes beyond disputes between owners and possessors. In crafting legal rules, it would be helpful to know, for instance, whether the attitudes and reactions of tenants (both short and long term) to property resemble or markedly differ from those of owners. Stern (2011) relies on empirical data to argue that owners and tenants of comparable residential duration perform

<sup>6</sup> This is particularly true for efficiency analysis, which aims to maximize people’s welfare as measured by the extent to which their preferences are satisfied (Shapiro and McClennen 1998). It also holds for objective theories of well-being (since they accept that fulfilling people’s wishes is one element of human welfare) and deontological theories (since they do not deny the importance of consequences, including human welfare). For a discussion of objective theories of well-being and their application to property law, see Lewinsohn-Zamir 2003.

quite similarly on certain citizenship measures such as community organizational participation and certain forms of collective action. Barros (2006, pp. 300–305) suggests that in many legal contexts there may be no reason to treat owners and renters differently. This is, in fact, the practice in Israel, where most lands are held in public ownership and individuals usually receive only long-term leases. Israeli law often equates between the rules applying to ownership and to leases for a term that exceeds twenty-five years. Moreover, recognizing that these tenants most probably regard themselves as owners and that it would be politically unfeasible to require them to vacate the land at the end of the lease, or else pay the full market value of a new lease, has led the state to gradually transfer ownership, in certain developed urban lands, to the long-term tenants (rather than to perpetually renew the lease without requiring payment from the tenants).

Regrettably, however, psychological studies usually do not differentiate between ownership and possession, and often conflate the two. Consequently, we cannot know whether their findings are attributable to ownership, possession, or both. Experimental studies of the EE constitute a good example of this difficulty. Typically, subjects are informed that they are owners of some object (such as a mug, a pen, or a chocolate bar), which they also physically possess. The value of the endowed object is elicited and often compared to the value of some alternative object, that subjects neither own nor possess (Knetsch and Sinden 1984; Kahneman, Knetsch, and Thaler 1990, 1991; Kahneman 1992; Morewedge et al. 2009; van Dijk and van Knippenberg 1996, 1998).<sup>7</sup> Terms pertaining to both ownership and possession are often used interchangeably throughout the study. In the relatively few experiments where ownership and possession were separated, it was usually not done in order to contrast the two, but rather for a different reason, such as to address the concern that by giving subjects possession of the endowed asset, the experimenters signal that it is more valuable than the alternative asset (Plott and Zeiler 2007, pp. 1455, 1459–60). The potential differences between lawful and unlawful possession were not investigated at all. Hence, to date, we do not really know to what extent the EE is an “ownership effect” or a “possession effect,” and what happens when the two work in opposition to each other.

These shortcomings of current behavioral research limit its applicability to legal issues, since nuances relevant for the law have not been tested yet. Notwithstanding, legal studies do apply such general data to owner-possessor disputes. A prominent example is the discussion on adverse possession. According to this doctrine, a trespasser may gain title to the land or immunity from an ejection suit, if her possession is adverse to the owner’s interest (i.e., does not stem from the owner’s right and is without the owner’s permission), actual, open and notorious, exclusive, and continuous for

<sup>7</sup> One notable exception is Knetsch and Wong (2009), who found an EE when subjects possessed—but were not yet the owners—of an endowed object. However, subjects did not own or possess the alternative object, and thus ownership and possession were not pitted against each other.

the statutory period of limitation. Stake (2001, pp. 2423–32), Cohen and Knetsch (1992, pp. 751–52) and Ellickson (1989, pp. 38–39) rely on the EE and loss aversion to justify adverse possession. The longtime possessor, so the argument goes, would perceive giving up the land as a loss, whereas the owner would view not receiving it back as a forgone gain. Since losses loom larger than gains, incurring the second cost—rather than the first—maximizes the parties’ joint welfare. Stake (2001, pp. 2459–64) also addresses the possibility that the nonpossessing owner may perceive the rejection of her suit as a loss (rather than an unattained gain), but believes that this loss would be smaller than the corresponding loss to the possessor. This is because the latter’s loss of a tangible physical object would be greater than the former’s loss of—what she probably sees as—only an intangible financial asset.

These claims seem plausible, yet without direct behavioral research it is difficult to know whether a nonpossessing owner would view herself as being in the domain of gains or in the domain of losses (Rose 2000, p. 489; Korobkin 2003, pp. 1259–62), or if she would perceive her land as merely a financial resource. Arguably, a registered landowner who loses her title or the ability of ever using the land would perceive this outcome as a loss of a nonfungible asset. Moreover, regardless of whether the situation is framed as an unattained gain or as a loss, a crucial factor is the magnitude of these gains and losses.<sup>8</sup> In real life, the owner may have purchased the land in the past for its market price and will therefore be deprived of this value, whereas the possessor may have knowingly trespassed on unpaid-for land. The value of the forgone gains or losses to the owner may greatly surpass the value of the losses to the possessor. In sum, one should be cautious in applying behavioral findings to legal issues absent a good “fit” between the two.

Note that even after the above issues are resolved, the case for adverse possession would not be settled one way or the other. The possible asymmetry in the subjective costs to each party—highlighted by behavioral studies—is but one relevant consideration among many (Ellickson 1989, p. 39 n. 42). Thus, for example, the high costs of adjudicating old cases sustained by all parties involved—plaintiffs, defendants, and courts—support the statute of limitation. However, in countries where adverse possession does not extinguish the owner’s title, acknowledgment of the possessor’s right to remain on the land would impair the accuracy of the information in the land registry, and the possibility of future purchasers to rely on it. For this reason, Israeli law, for example, does not recognize adverse possession claims in lands that have undergone a settlement of title procedure (section 159(b) of the Land Law, 5729-1969, 23 LSI 283, 311).

<sup>8</sup> To the extent that the possessor has improved the land, she may be reimbursed for these costs, at least in cases where the encroachment was in good faith. See Sterk 1987, pp. 80–81; *Somerville v. Jacobs*, 170 S.E.2d 805, 807, 813 (1969).

### 3.4 Property Rules versus Liability Rules

Behavioral studies offer relevant considerations for choosing the remedy for rights-infringement, including property rights. Following the influential article by Guido Calabresi and Douglas Melamed (1972), this issue is known as the choice between property rules and liability rules. An entitlement is protected by a “property rule” if no one can appropriate the entitlement without securing the owner’s consent. The entitlement must be transferred through a voluntary transaction, with the owner-seller agreeing to its price. In contrast, liability rule protection enables a forced transfer of the entitlement. The coercing party need not seek the owner’s consent, but must only pay her the objectively determined value of the entitlement. Calabresi and Melamed proposed an elegant criterion for choosing between the two forms of entitlement-protection—transaction costs. Property rules should be employed when transaction costs are low and the parties can bargain to achieve desirable outcomes. Liability rules should be used when transaction costs are high, such as when numerous parties are involved (Calabresi and Melamed 1972, pp. 1092–93, 1105–10, 1118–19, 1125–27).

Although the Calabresi-Melamed criterion has garnered a great deal of support (e.g., Craswell 1993, pp. 8–9, 15; Krier and Schwab 1995, pp. 450–51; Merges 1994), it has also been criticized. Scholars have claimed that liability rules may be superior to property rules even when transaction costs are low. This argument emphasizes the risk that bargaining under property rules might fail. This can happen, for example, in a bilateral monopoly situation with asymmetric information. If each party attempts to capture all the gains from the trade and miscalculates the other party’s evaluations and reactions, then potentially efficient transactions might not take place at all, or only after a costly delay. An advantage of liability rules in this regard is that they remove the entitlement owner’s holdout power, thereby ensuring the execution of efficient transfers (Ayres and Talley 1995, pp. 1030–33, 1042–44, 1055–56; Kaplow and Shavell 1996, pp. 717–18, 724–37; Shavell 2004, pp. 87–91, 315).

However, the employment of liability rules is not without risk, since courts might err in assessing the entitlement owner’s losses. Even advocates of liability-rule protection concede that the argument in favor of liability rules holds only if courts do not systematically underestimate compensation awards to entitlement owners (Kaplow and Shavell 1996, pp. 720, 730–32). Thus, the superiority of liability rules over property rules rests on the presumption that the risks involved in employing the latter are higher than those involved in employing the former.

Some scholars have claimed that the EE supports this crucial assumption. Thus, Buccafusco and Sprigman (2010, 2011) argue that intellectual property rights produce in creators both an EE (as a result of ownership) and an additional “creativity effect” (due to being the creators of the owned work), which make them overly reluctant to part with their work (see also Jolls and Sunstein 2006, pp. 220–22; Lemley 2012, pp. 485–86. For a critique of Buccafusco and Sprigman’s argument, see Tur-Sinai 2011).

In a similar vein, Penalver and Katyal (2007, pp. 1134–45) assert that the EE creates a conservative bias in favor of current property owners that hinders beneficial social changes. Likewise, Ayres (1998, pp. 810–12) generally states that property rule protection of entitlements creates an EE that impedes negotiations. According to this literature, liability rule protection weakens or eliminates the EE, because entitling owners only to monetary damages conveys a lesser sense of ownership than entitling them to an injunction against interference with their property (Rachlinski and Jourden 1998).

Notwithstanding, one may argue that, overall, the existing behavioral data are more supportive of the opposite argument, namely, that property rules are the superior remedy when transaction costs are low and parties can bargain with one another. Generally speaking, psychological studies invite more optimism about people's ability to reach a mutual agreement, and suggest that miscalculations under liability rules may be a graver danger than presently realized.

The fear that voluntary transactions under property-rule protection might fail rests on an assumption of extreme greediness (each party strives to capture the entire surplus of the transaction). Yet an extensive body of research challenges this presumption. Individuals do not always exercise holdout power, and frequently cooperate with one another—and thus succeed in fairly dividing the potential gains from trade (Etzioni 1988; Jolls, Sunstein, and Thaler 1998; Kahneman, Knetsch, and Thaler 1986; Korobkin and Ulen 2000, pp. 1135–38, 1138–41; Mansbridge 1990; Margolis 1982). Of particular interest is the literature on the Ultimatum Game. In the basic form of this game, one person (the proposer) is asked to divide a sum of money between herself and another person (the responder). The responder is free to accept or reject the sum offered. In the former case, the proposed division is implemented. In the latter, neither player receives anything. The standard economic prediction is one of utmost opportunism: the proposer will offer the smallest unit of money possible in the game and the responder will accept this sum, since any share of the pie is preferable to receiving nothing. Experimental results deviate dramatically from this prediction. Most proposers offer substantial amounts of money, often an equal split of the pie, and responders accept generous offers. Offers that deviate substantially from 50 percent (such as less than 20 percent) are typically rejected. The rejection of low offers is explained both by resistance to unfairness and by the desire to punish ungenerous proposers (for overviews of the vast literature on ultimatum games, see Hoffman, McCabe, and Smith 2008, pp. 411–28, 436–53; Kagel and Roth 1995, pp. 253–348; Thaler 1992, pp. 21–35).

In the context of negotiations for the transfer of an entitlement, the responder in an ultimatum game may be compared to an owner whose entitlement is protected by a property rule. She may agree to an offer to sell her entitlement for the price suggested by the proposer-buyer, which represents a way of splitting the gains from the trade. If the owner rejects the buyer's proposal, neither party would realize any gains. The results of ultimatum games question the conventional assumption that property-rule protection induces sellers to hold out for all the profits from the trade. Most ultimatum bargaining is successful and a minority of offers are rejected for being too low. Proposers know in advance that they must acquire the responders' consent to enjoy any gains.

Consequently, most proposers make offers that they expect responders to accept, and their offers are indeed accepted and efficient transactions carried out. The rejection power afforded to responders does not induce extreme greediness, but only prevents proposers from trying to claim all the profits for themselves. Ultimatum experiments thus attest to the success of even one-period bargaining between complete strangers, who interact anonymously and are not troubled by considerations of long-term relationships, their reputation, and so forth. Negotiations in real life would plausibly be at least equally successful, because bargainers communicate directly with one another and have ample opportunity to discuss their disagreements and negotiate an acceptable compromise (for a more extensive treatment of these ideas, see Lewinsohn-Zamir 2001a).<sup>9</sup> Furthermore, the argument that negotiations would fail under a property rule assumes that sellers are risk-seeking. However, some behavioral studies have found that sellers tend to be risk-averse (i.e., more concessionary and willing to compromise), while buyers tend to be risk-seeking (i.e., holding out longer for a higher future payoff) (Neale and Bazerman 1991, pp. 156–57; Neale, Huber, and Northcraft 1987).

The use of the EE to strengthen the argument in favor of liability rules is particularly problematic in the context of property law. If the very fact of having property creates a strong EE that obstructs efficient transactions, and if the solution is to protect property rights with liability rules, wouldn't this undermine the whole idea of private property? The right of property owners to exclude others from using the property without their consent is considered a vital component of ownership (e.g., Merrill 1998). Adopting a liability rule regime would radically alter the current institution of private property, with detrimental effects in terms of efficiency and fairness.

Parenthetically, such radical restructuring of property may account for the results of Rachlinski and Jourden (1998). This experimental study found that a stronger EE is manifested when entitlements are protected by an injunction (property rule) than when they are protected by a damages remedy (liability rule) (pp. 1559–72). The authors compared the sale prices of entitlements protected *solely* by liability rules to the sale prices of entitlements protected only with injunctions. Both regimes greatly differ from any property institution in the real world. In addition, all questionnaires informed the subjects that if they agreed to sell the entitlement, the purchase price could be used for acquiring a good substitute, and that this alternative goal could not be realized without a sale. The existence of an adequate substitute may have reduced subjects' EE in the liability-rule scenario, because the loss caused by the possibility of a forced transfer could be immediately remedied by a close substitute. In contrast, subjects in the property rule scenario may have felt that a sale was redundant, as they could achieve a similar goal simply by holding onto their current entitlement (for further discussion of this

<sup>9</sup> Note that bargaining under a property rule will usually be carried out in advance—that is to say, before any breach of the entitlement owner's rights has occurred. Therefore, we should not fear the same degree of animosity that can obstruct successful bargaining conducted *ex post* (i.e., after such a breach), before the filing of a lawsuit, or around an existing court order. On these problems, see Jolls, Sunstein, and Thaler (1998, pp. 1497–98), and Farnsworth (1999).



study see Lewinsohn-Zamir 2001a, p. 254 n. 121; Korobkin 2003, pp. 1284–85 [explaining why the positive implications of the study might be limited due to its environmental context]).

At any rate, the EE does not seem to call for a drastic revision in our understanding of private property. This is because the existence and magnitude of the EE vary from one case to another. The argument in favor of liability rules relies on the general EE phenomenon. Yet the circumstances in which it is reduced or eliminated are highly relevant for the current debate. Studies have shown, for instance, that goods held for exchange—in contrast to goods held for use—do not produce an EE, nor do money, bargaining chips, vouchers, or tokens that are valued only for their trading possibilities (Hoffman and Spitzer 1993, pp. 78–82, 111, 113; Kahneman, Knetsch, and Thaler 1990, pp. 1328–32, 1344; Novemsky and Kahneman 2005b; van Dijk and van Knippenberg 1996).<sup>10</sup> In a similar vein, the EE is substantially smaller when an adequate substitute for the relevant good is available (Adamowicz, Bhardwaj, and Macnab 1993).<sup>11</sup> Contrarily, the more difficult it is to compare the endowed item and its alternative, the greater the reluctance to trade (van Dijk and van Knippenberg 1998).

Furthermore, experimental research has found that the EE is significantly weakened or eliminated when an owner *wants* to sell her asset. Thus, Mandel (2002) demonstrated that the motivation to bring about a potential transaction—which he labels “transaction demand”—moderates the EE. As transaction demand increases, owners are more inclined to sell at lower prices and prospective buyers are more willing to pay higher prices. In addition, when transaction demand was high for both sellers and buyers, the EE was reversed. Indeed, if transaction demand did not exist in the real world and if “motivational factors did not often override the effects of loss aversion on valuation, then far fewer transactions would be expected to occur than they actually do” (Mandel 2002, p. 745). Similarly, Simonson and Drolet (2004) showed that if a person decides that she wants to sell an item, no EE exists and the market price becomes the primary driver of the minimum asking price. In contrast, if a person is uncertain about her desire to trade, an EE is created.

All these findings support the argument that the use of property rules may produce a weaker EE than the use of liability rules. Property rule protection means that an owner will sell only if she has reached the stage at which she wishes to sell. A person desiring to sell is a person who intends to participate in the market and who believes that good substitutes exist for her entitlement. Put differently, property rules may induce a frame of mind that focuses on the profits of the exchange—the money or alternative good that

<sup>10</sup> However, an EE exists if the value of the financial instruments or bargaining chips is uncertain. See Samuelson and Zeckhauser 1988, pp. 12–22; van Dijk and van Knippenberg 1996.

<sup>11</sup> The authors found that differences between selling and buying prices for tickets to a National Hockey League game considerably decreased when a substitute for personal attendance—live television and radio broadcasts of the game—existed. The claim that adequate substitutes decrease the magnitude of the EE is argued for theoretically in Hanemann 1991.



can be bought with it—rather than on the good that is given up.<sup>12</sup> Liability rule protection, in contrast, means that an owner can be forced to transfer her entitlement. Even if the parties attempt to negotiate a transfer of the entitlement in the shadow of the liability rule, the owner is aware that refusal to sell will be of no avail. In any case, the entitlement can be taken against her will. The power of coercion might evoke the role of an “unwilling seller” and an attitude of resistance. An individual unwilling to sell is an individual who prefers to keep the entitlement and who believes that adequate substitutes for the entitlement do not exist. Consequently, the EE might be stronger when entitlements are protected with liability rules (Lewinsohn-Zamir 2001a).

Studies on the connection between possessions and self-identity, self-image, and self-development may also be relevant for the property rule–liability rule debate. Beggan (1992) found that overvaluation of owned objects occurs particularly when owners’ positive self-image is threatened (as a result, e.g., of ego-injuring negative performance feedback). Belk (1988) showed that involuntary loss of possessions causes trauma and loss of self, but voluntary disposition does not. If forced transfer of assets through liability rules injures owners’ self-image, this might increase their valuation of the taken assets and the loss that they would suffer. It is plausible that voluntary transfer through property rule protection would not cause these detrimental effects (at all, or to a lesser extent).

Once a transfer is coerced under a liability rule, the owner is entitled to compensation for her losses. Behavioral studies suggest that quantifying the damages award is a more complex task than commonly assumed. As explained earlier, there is evidence that people prefer in-kind remedies over monetary remedies, and that this preference exists even with respect to fungible property, easily replaceable in the market (Lewinsohn-Zamir 2013). Accordingly, all assets are to some extent unique, in the sense that it is difficult to quantify the loss incurred by receiving a monetary relief rather than an in-kind one. Furthermore, it appears that people perceive outcomes broadly, and therefore judge events according to several variables beyond their end results. For example, a factor that exacerbates an injurious outcome is whether the injury was brought about intentionally (though without malice or intent to harm) rather than unintentionally. Another factor of relevance is the voluntariness or nonvoluntariness of an act. Unwilling transfer of even exchange goods inflicts on the owner a loss that is distinct from her subjective valuation of the good (Lewinsohn-Zamir 2012). Since exercise of liability rules entails intentional injury, coerced transfer, and monetary redress,

<sup>12</sup> Novemsky and Kahneman (2005a, p. 140) agree that people’s intentions may moderate loss aversion: “when there is an intention to exchange a good, the reference point is not the current endowment but rather the expected endowment at the conclusion of the exchange. In this case, giving up a good in an exchange would not involve loss aversion, because the good is not part of the reference point (i.e., the expected endowment).” See also Liberman et al. (1999), who found that subjects were reluctant to exchange an endowed good when they were in a “prevention” frame of mind (caused by priming that was unrelated to the subsequent exchange experiment), but not when they were in a “promotion” frame of mind.

there is reason to fear that the prevailing compensation criterion of “fair market value” may be inherently undercompensatory, and might lead in the long run to inefficiency and demoralization of entitlement owners (Lewinsohn-Zamir 1996, pp. 61–62, 65–66). An advantage of property rules in this regard is that they bypass the need for complicated damages calculations.

In sum, further research is needed in order to clarify the behavioral effects of employing property rules and liability rules. Until then, one cannot assume—as standard economic analysis does—that failed negotiations for voluntary transfers under property rule protection are a greater risk than undercompensating entitlement owners under liability rule protection.

### 3.5 Redistribution

Most people agree that promoting equality in society is an important goal of the state. Much more controversial, however, is the question which means should be used to redistribute welfare. Specifically, an ongoing debate has centered on whether redistribution should be attained solely through taxes and transfer payments (such as progressive taxation, negative taxes, unemployment compensation, and cash assistance to needy families), or also via private law, including substantive rules of property, tort, and contract law (hereinafter, legal rules). Law-and-economics scholars have argued that legal rules should not be used for redistribution, because they are more costly and less effective than the tax-and-transfer alternative (e.g., Cooter and Ulen 2012, pp. 7–8, 106–8). When legal rules redistribute income in favor of the poor, so the argument goes, they distort people’s work incentives just as much as the tax system does. People would respond to the redistributive legal rule in the same way they respond to an increase in their marginal tax rates, and may consequently choose leisure over labor. Legal rules, however, create an additional inefficiency—the distortion in the very behavior that the legal rules aimed to regulate. For instance, a 30 percent marginal tax rate together with an inefficient tort rule that redistributes 1 percent of wealthy defendants’ income to poor plaintiffs would distort work incentives to the same extent as a 31 percent tax rate coupled with an efficient tort rule. The former regime, however, entails the additional costs involved in defendants taking excessive precautions and refraining from efficient activities (Kaplow and Shavell 1994, 2000; Polinsky 2003, pp. 148–49, 153, 155). In addition, legal rules are less effective than taxes and transfers, that is, they achieve less of the desired distributive outcomes. This is because in contractual settings (as opposed to circumstances in which bargaining is impractical) the market often responds in a way that wholly or partially offsets the redistribution (Kaplow and Shavell 2001, pp. 993, 1126; Weisbach 2003, pp. 448–49). Take, for example, a mandatory quality standard in favor of tenants, requiring landlords to lease residential units that are fit for human habitation (on the implied warranty of habitability, see Stoebuck and Whitman (2000), sections 6.38–40). Increasing landlords’ costs is liable to increase rents and reduce

the supply of low-rent housing, thereby harming the poorest tenants (Posner 2011, pp. 645–48).

Behavioral studies, however, strengthen the case for redistributing through legal rules as well. Jolls, for example, relied on the robust finding that people are often unrealistically optimistic and underestimate the probability of negative events (see Baker and Emery 1993; Weinstein [1980, 1987]). Taxation is a certain event, whereas an event triggering the application of a redistributive legal rule—such as being involved in a car accident—is an uncertain one. If potential defendants underestimate the latter risk, they will perceive its costs to be lower than they actually are. Consequently, a redistributive tort rule will distort their work incentives less than a tax yielding the same amount of revenue for the government (Jolls 1998, pp. 1658–63).

Importantly for our purposes, behavioral findings question one of the basic premises of the economic critique of redistributive legal rules. The economic argument assumes that the success of the redistribution should be evaluated according to the bare quantity of resources that people receive, regardless of the way they were obtained. That is to say, the method generating the distributive outcome is irrelevant and does not affect its goodness. However, numerous experimental studies have shown that the benefit people derive from resources depends on complex factors, including the acts that generate the resources and the source from which they are received. Thus, for example, an object attained “as of right,” through effort or as a result of success, is valued much more highly than a similar object obtained through no entitlement, chance, or due to failure (Loewenstein and Issacharoff 1994; Hoffman et al. 1994). Likewise, the identity of the person who gave an asset affects its valuation and subsequent use by the recipient (McGraw, Tetlock, and Kristel 2003). These phenomena apply even to money (Zelizer 1997, pp. 3, 5, 200, 209, 211; Thaler 2000, p. 259). In legal contexts, it was found that factors such as how an outcome was brought about (with goodwill and cooperation or not), the identity of the parties involved (strangers or friends), the voluntariness or nonvoluntariness of their behavior, and the intentionality or nonintentionality of their acts, significantly affect its valuation by both laypersons and businesspeople (Lewinsohn-Zamir 2012). In a similar vein, studies of procedural justice have demonstrated that people care not only about substantive outcomes, but also about the process leading up to them. The perceived fairness or unfairness of the process may affect how the outcome is accepted and how legitimate it is regarded to be (Thibaut and Walker 1975, pp. 72–77, 89–94, 118–22; Tyler 1989).

This body of research indicates that some methods of redistribution are likely to be perceived as humiliating, whereas others would plausibly be regarded as more respectful and empowering. Taxes and transfers are often viewed as “charity giving” (e.g., Kluegel and Smith 1986, pp. 152–57, 163–65, 175, 293), whereas redistributive legal rules set the baseline for interactions between individuals and typically convey a message of entitlement (see also Singer 2000, p. 177). Consequently, a smaller quantity of goods obtained through legal rules may advance the recipients’ welfare to a greater extent than the same or even larger amount received through taxes and transfer payments (Lewinsohn-Zamir 2006). The implied warranty of habitability in landlord and

tenant law, for instance, carries with it a message that the object of a lease can only be a house fit for human habitation. This rule applies to all tenants, not only to poor ones. Accordingly, tenants have a right to habitable housing, and a landlord complying with this requirement is simply fulfilling her obligation, not granting any favors (for further treatment of these ideas see Lewinsohn-Zamir 2006).<sup>13</sup>

The argument that the method of redistribution is important can be further demonstrated with the choice of a marital property system. The joint property regime holds that property earned by either spouse during the marriage is owned jointly and equally by both spouses, notwithstanding formal title. Each spouse has similar, vested property rights in the marital assets, which can be exercised at any point during the course of the relationship. The separate property regime, in contrast, maintains that each spouse separately retains the property that he or she earned. A more equitable division of assets is attained upon divorce, according to factors like need, contribution in the home, health, and occupation. In addition, the poorer spouse may be awarded alimony or maintenance payments (Singer 2005, pp. 397, 402–11).

The behavioral findings lend support for the superiority of the joint property regime as a redistributive device. Alimony and maintenance payments (like taxes and transfer payments) resemble charity giving. They imply that the spouse has no right to the property accumulated throughout the marriage, but rather only to receiving some of the other spouse's money, in order to make ends meet. This type of redistribution may be associated with failure, thus diminishing the value of the thing received. Although in-kind transfer of assets upon divorce is superior to cash payments, it too might convey a message of "no entitlement," because it is likely to be perceived as a transfer of another person's property. The fact that the extent of the transfer is determined by factors like needs, rehabilitation, health, and skills may also solidify the association of the redistribution with handouts to the needy. In contrast, joint ownership grants both spouses equal standing and rights in the marital assets throughout their relationship, and therefore may foster notions of entitlement, responsibility and sharing. In addition, having assets "as of right" and due to effort is more strongly associated with success (Lewinsohn-Zamir 2006, pp. 385–89). Once again, it would be helpful if future behavioral research would directly test these hypotheses. Nevertheless, the existing data cast serious doubt on the standard economic assumption that redistribution is simply a quantitative matter.

<sup>13</sup> Even if the implied warranty of habitability fails to redistribute income—because the market responds by raising the rents or altering other terms in the lease, it may still succeed in redistributing the objective good of minimal quality housing (see Lewinsohn-Zamir 2006, pp. 340–51). Furthermore, some writers argue that in certain circumstances, the implied warranty of habitability would successfully transfer wealth to tenants. See Ackerman 1971, pp. 1097–98, 1102–19, 1186–88; Kennedy 1987, pp. 497–506.

## 4 CONCLUDING REMARKS

---

This chapter has critically surveyed the behavioral literature on property, and in doing so discussed several psychological findings that are relevant for property law. A question that commonly arises with respect to such findings is whether the law should attempt to “debias” people and steer them in a more rational direction (for general discussions of debiasing through law, see Blumenthal 2007; Jolls and Sunstein 2006; chapter 6 by Pi, Parisi, and Luppi in this volume and chapter 28 by Sunstein in this volume). The answer to this question is necessarily context dependent. Sometimes, there is actually no justification for legal intervention. For instance, if people largely prefer in-kind remedies over monetary ones, there is no compelling reason to try to change these preferences, since there is nothing objectionable or irrational about them. In other cases, an attempt to debias may be futile, as in the case of people’s tendency to regard events with similar end results as generating different outcomes (Lewinsohn-Zamir 2012). It is probable that people cannot be taught to equate outcomes with end results, due to the relative rareness of factors that adversely affect their valuation of events. Take, for example, ownership of land. Most parcels are voluntarily sold in the market and the great majority of landowners never have their land condemned. Consequently, it is not surprising if extraordinary features of an event, like the coerciveness of a transfer, affect the valuation of the outcome and increase the landowner’s losses. Since expropriations are relatively uncommon, it is unlikely that the law can educate people that nonvoluntary purchases are not substantively different from voluntary ones, and that they should not expect any extra recompense for being forced to part with their land.

A different obstacle in the way of debiasing is that the same bias may be negative in certain circumstances, but positive in others. Overoptimism, for example, is injurious if it makes borrowers underestimate the likelihood of sharp fluctuations in income that would cause them to go bankrupt and lose most of their property (Block-Lieb and Janger 2006, pp. 1540–43). Contrarily, in intellectual property contexts, overoptimism is regarded as beneficial, since it leads authors and inventors to assume risks and invest time, money and effort in creative activities (Tur-Sinai 2011, pp. 154–56). It may not be possible to debias the “bad” overoptimism while retaining the “good.” Consequently, case-specific mandatory regulation may be necessary, such as bankruptcy rules that exempt some of the debtor’s property from the reach of her creditors (Lewinsohn-Zamir 2003, pp. 1722–30).

In closing, a comment on the future of the behavioral approach to property law is in order.

Undoubtedly, the existing body of research has contributed significantly to diverse issues and debates in the field. At the same time, there are ways in which the behavioral approach can be improved so as to increase its relevance and fruitfulness for property law. Two main avenues for improvement come to mind.

First, scholars should aim to broaden the scope of behavioral studies. It would be worthwhile to cover additional property topics, rather than continue concentrating on land, and specifically on residences. Further research should focus on movables and intangible property, as well as on property used for commercial purposes. In a similar vein, current behavioral studies mainly test the efficiency and personhood theories of property, while largely ignoring others, such as the labor and libertarian theories. Furthermore, from the numerous behavioral phenomena established in psychological studies, jurists have mostly employed the endowment effect. While this makes sense when discussing property law, future research should make more use of additional behavioral insights and biases. For instance, the literature on cognitive dissonance (see Festinger 1957; Cooper 2007) and on intrinsic and extrinsic motivation (see Aronson 2008; Kelman 1961) may have fruitful implications for people's compliance with rules of property law (see also Lewinsohn-Zamir 2014).

Second, there is room for enhancing the "fit" between behavioral studies and the legal issues they address. Legal scholars should design experiments that are more carefully tailored to property topics. For example, distinctions between ownership and possession, between lawful and unlawful possession, and between prospective and actual possession, may be of little importance to psychologists but are highly relevant for legal policymaking.

## REFERENCES

- Ackerman, B. A. 1971. Regulating Slum Housing Markets on Behalf of the Poor: Of Housing Codes, Housing Subsidies and Income Redistribution Policy. *Yale Law Journal* 80:1093–197.
- Ackerman, B. A. 1977. *Private Property and the Constitution*. New Haven: Yale University Press.
- Adamowicz, W. L., V. Bhardwaj, and B. Macnab. 1993. Experiments on the Difference Between Willingness to Pay and Willingness to Accept. *Land Economics* 69:416–27.
- Alterman, R. 2010. *Takings International: A Comparative Perspective on Land Use Regulations and Compensation Rights*. Chicago: American Bar Association Publishing.
- Aronson, E. 2008. *The Social Animal*. 10th ed. New York: Worth Publishers.
- Ayres, I., and E. Talley. 1995. Solomonian Bargaining: Dividing a Legal Entitlement to Facilitate Coasean Trade. *Yale Law Journal* 104:1027–117.
- Ayres, I. 1998. Protecting Property with Puts. *Valparaiso Law Review* 32:793–831.
- Baker, L.A., and R. E. Emery. 1993. When Every Relationship Is above Average: Perceptions and Expectations of Divorce at the Time of Marriage. *Law and Human Behavior* 17:439–50.
- Baron, J., and I. Ritov. 1994. Reference Points and Omission Bias. *Organizational Behavior and Human Decision Processes* 59:475–98.
- Barros, D. B. 2006. Home as a Legal Concept. *Santa Clara Law Review* 46:255–306.
- Barros, D. B. 2009. Legal Questions for the Psychology of Home. *Tulane Law Review* 83:645–60.
- Beggan, J. K. 1992. On the Social Nature of Nonsocial Perception: The Mere Ownership Effect. *Journal of Personality and Social Psychology* 62:229–37.
- Belk, R. W. 1988. Possessions and the Extended Self. *Journal of Consumer Research* 15:139–68.



- Block-Lieb, S., and E. J. Janger. 2006. The Myth of the Rational Borrower: Rationality, Behavioralism, and the Misguided “Reform” of Bankruptcy Law. *Texas Law Review* 84:1481–566.
- Blumenthal, J. A. 2007. Emotional Paternalism. *Florida State University Law Review* 35:1–72.
- Blumenthal, J. A. 2009. “To Be Human”: A Psychological Perspective on Property Law. *Tulane Law Review* 83:609–44.
- Buccafusco, C., and C. Sprigman. 2010. Valuing Intellectual Property: An Experiment. *Cornell Law Review* 96:1–45.
- Buccafusco, C., and C. Sprigman. 2011. The Creativity Effect. *University of Chicago Law Review* 78:31–52.
- Calabresi, G., and A. D. Melamed. 1972. Property Rules, Liability Rules, and Inalienability: One View of the Cathedral. *Harvard Law Review* 85:1089–128.
- Camerer, C. 1995. Individual Decision Making. In *The Handbook of Experimental Economics*, ed. J. H. Kagel and A. E. Roth, 587–703. Princeton, NJ: Princeton University Press.
- Cohen, D., and J. L. Knetsch. 1992. Judicial Choice and Disparities between Measures of Economic Values. *Osgoode Hall Law Journal* 30:737–70.
- Cooper, J. 2007. *Cognitive Dissonance: Fifty Years of a Classic Theory*. London: Sage Publications.
- Cooter, R., and T. Ulen. 2012. *Law and Economics*. 6th ed. Boston: Pearson Addison Wesley Publishing.
- Craswell, R. 1993. Property Rules and Liability Rules in Unconscionability and Related Doctrines. *University of Chicago Law Review* 60:1–66.
- Csikszentmihalyi, M., and E. Rochberg-Halton. 1981. *The Meaning of Things: Domestic Symbols and the Self*. Cambridge: Cambridge University Press.
- Dagan, H. 2011. *Property: Values and Institutions*. New York: Oxford University Press.
- Ellickson, R. C. 1989. Bringing Culture and Human Frailty to Rational Actors: A Critique of Classical Law and Economics. *Chicago-Kent Law Review* 65:23–56.
- Etzioni, A. 1988. *The Moral Dimension: Toward a New Economics*. New York: Free Press.
- Farnsworth, W. 1999. Do Parties to Nuisance Cases Bargain after Judgment? A Glimpse Inside the Cathedral. *University of Chicago Law Review* 66:373–436.
- Fennell, L. A. 2003. Death, Taxes, and Cognition. *North Carolina Law Review* 81:567–652.
- Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Evanston, IL: Row, Peterson.
- Fischel, W. A. 1995. *Regulatory Takings: Law, Economics, and Politics*. Cambridge, MA: Harvard University Press.
- Gneezy, U., and A. Rustichini. 2000. A Fine Is a Price. *Journal of Legal Studies* 29:1–18.
- Godsil, R. D. 2004. Viewing the Cathedral from behind the Color Line: Property Rules, Liability Rules, and Environmental Racism. *Emory Law Journal* 53:1807–85.
- Hanemann, W. M. 1991. Willingness to Pay and Willingness to Accept: How Much Can They Differ? *American Economic Review* 81:635–47.
- Hoffman, E., and M. Spitzer. 1993. Willingness to Pay vs. Willingness to Accept: Legal and Economic Implications. *Washington University Law Review* 71:59–114.
- Hoffman, E., K. McCabe, K. Shachat, and V. Smith. 1994. Preferences, Property Rights, and Anonymity in Bargaining Games. *Games and Economic Behavior* 7:346–80.
- Hoffman, E., K. McCabe, and V. Smith. 2008. Reciprocity in Ultimatum and Dictator Games: An Introduction. In *Handbook of Experimental Economics Results*, vol. 1, ed. C. R. Plott and V. L. Smith, 411–53. Amsterdam: North-Holland.



- Isoni, A., G. Loomes, and R. Sugden. 2011. The Willingness to Pay–Willingness to Accept Gap, the “Endowment Effect,” Subject Misconceptions, and Experimental Procedures for Eliciting Valuations: Comment. *American Economic Review* 101:991–1011.
- Jolls, C. 1998. Behavioral Economics Analysis of Redistributive Legal Rules. *Vanderbilt Law Review* 51:1653–78.
- Jolls, C., and C. R. Sunstein. 2006. Debiasing through Law. *Journal of Legal Studies* 35:199–241.
- Jolls, C., C. R. Sunstein, and R. H. Thaler. 1998. A Behavioral Approach to Law and Economics. *Stanford Law Review* 50:1471–550.
- Kahneman, D. 1992. Reference Points, Anchors, Norms, and Mixed Feelings. *Organizational Behavior and Human Decision Processes* 51:296–312.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1986. Fairness and the Assumptions of Economics. In *Rational Choice: The Contrast between Economics and Psychology*, ed. R. M. Reder and R. M. Hogarth, 101–16. Chicago: University of Chicago Press
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1990. Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy* 98:1325–48.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1991. The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives* 5:193–206.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 2008. The Endowment Effect: Evidence of Losses Valued More Than Gains. In *Handbook of Experimental Economics Results*, vol. 1, ed. C. R. Plott and V. L. Smith, 939–55. Amsterdam: North-Holland.
- Kaplow, L., and S. Shavell. 1994. Why the Legal System Is Less Efficient Than the Tax in Redistributing Income. *Journal of Legal Studies* 23:667–82.
- Kaplow, L., and S. Shavell. 1996. Property Rules versus Liability Rules: An Economic Analysis. *Harvard Law Review* 109:713–90.
- Kaplow, L., and S. Shavell. 2000. Should Legal Rules Favor the Poor? Clarifying the Role of Legal Rules and the Income Tax in Redistributing Income. *Journal of Legal Studies* 29:821–36.
- Kaplow, L., and S. Shavell. 2001. Fairness versus Welfare. *Harvard Law Review* 114:961–1389.
- Kelly, J. J. 2006. “We Shall Not Be Moved:” Urban Communities, Eminent Domain and the Socioeconomics of Just Compensation. *St. John’s Law Review* 80:923–90.
- Kelman, H. C. 1961. Processes of Opinion Change. *Public Opinion Quarterly* 25:57–78.
- Kennedy, D. 1987. The Effect of the Warranty of Habitability on Low Income Housing: “Milking” and Class Violence. *Florida State University Law Review* 15:485–520.
- Kluegel, J. R., and E. R. Smith. 1986. *Beliefs about Inequality: Americans’ Views of What Is and What Ought to Be*. New York: A. de Gruyter.
- Knetsch, J. L., and J. A. Sinden. 1984. Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value. *Quarterly Journal of Economics* 99:507–21.
- Knetsch, J. L., and W. Wong. 2009. The Endowment Effect and the Reference State: Evidence and Manipulations. *Journal of Economic Behavior and Organization* 71:407–13.
- Korobkin, R. B. 2003. The Endowment Effect and Legal Analysis. *Northwestern University Law Review* 97:1227–93.
- Korobkin, R. B., and T. S. Ulen. 2000. Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics. *California Law Review* 88:1051–144.
- Krier, J. E., and S. J. Schwab. 1995. Property Rules and Liability Rules: The Cathedral in Another Light. *New York University Law Review* 70:440–84.
- Lemley, M. A. 2012. Contracting around Liability Rules. *California Law Review* 100:463–86.

- Lewinsohn-Zamir, D. 1996. Compensation for Injuries to Land Caused by Planning Authorities: Towards a Comprehensive Theory. *University of Toronto Law Journal* 46:47–127.
- Lewinsohn-Zamir, D. 2001a. The Choice between Property Rules and Liability Rules Revisited: Critical Observations from Behavioral Studies. *Texas Law Review* 80:219–60.
- Lewinsohn-Zamir, D. 2001b. Contemporary Property Law Scholarship: A Comment. *Theoretical Inquiries in Law* 2:97–105.
- Lewinsohn-Zamir, D. 2003. The Objectivity of Well-Being and the Objectives of Property Law. *New York University Law Review* 78:1669–754.
- Lewinsohn-Zamir, D. 2006. In Defense of Redistribution in Private Law. *Minnesota Law Review* 91:326–97.
- Lewinsohn-Zamir, D. 2008. More Is Not Better Than Less: An Exploration in Property Law. *Minnesota Law Review* 92:634–713.
- Lewinsohn-Zamir, D. 2012. Taking Outcomes Seriously. *Utah Law Review* 2012:861–902.
- Lewinsohn-Zamir, D. 2013. Can't Buy Me Love: Monetary versus In-Kind Remedies. *University of Illinois Law Review* 2013:151–94.
- Lewinsohn-Zamir, D. 2014. Do the Right Thing: Indirect Remedies in Private Law. *Boston University Law Review* 94:55–103.
- Lieberman, N., L. C. Idson, C. J. Camacho, and E. T. Higgins. 1999. Promotion and Prevention Choices between Stability and Change. *Journal of Personality and Social Psychology* 77:1135–45.
- Loewenstein, G., and S. Issacharoff. 1994. Source Dependence in the Valuation of Objects. *Journal of Behavioral Decision Making* 7:157–68.
- Mandel, D. R. 2002. Beyond Mere Ownership: Transaction Demand as a Moderator of the Endowment Effect. *Organizational Behavior and Human Decision Processes* 88:737–47.
- Mansbridge, J. J. 1990. *Beyond Self-Interest*. Chicago: University of Chicago Press.
- Margolis, H. 1982. *Selfishness, Altruism and Rationality: A Theory of Social Choice*. Cambridge: Cambridge University Press.
- McCaffery, E. J. 1994. Cognitive Theory and Tax. *UCLA Law Review* 41:1861–947.
- McGraw, A. P., P. E. Tetlock, and O. V. Kristel. 2003. The Limits of Fungibility: Relational Schemata and the Value of Things. *Journal of Consumer Research* 30:219–29.
- Merges, R. P. 1994. Of Property Rules, Coase, and Intellectual Property. *Columbia Law Review* 94:2655–73.
- Merrill, T. W. 1998. Property and the Right to Exclude. *Nebraska Law Review* 77:730–55.
- Merrill, T. W., and H. E. Smith. 2007. The Morality of Property. *William and Mary Law Review* 48:1849–96.
- Michelman, F. I. 1967. Property, Utility and Fairness: Comments on the Ethical Foundations of “Just Compensation” Law. *Harvard Law Review* 80:1165–258.
- Michelman, F. I. 1988. Takings, 1987. *Columbia Law Review* 88:1600–29.
- Morewedge, C. K., L. L. Shu, D. T. Gilbert, T. D. Wilson. 2009. Bad Riddance or Good Rubbish? Ownership and Not Loss Aversion Causes the Endowment Effect. *Journal of Experimental Social Psychology* 45:947–51.
- Nadler, J., and S. Diamond. 2008. Eminent Domain and the Psychology of Property Rights: Proposed Use, Subjective Attachment, and Taker Identity. *Journal of Empirical Legal Studies* 5:713–49.
- Nash, J. R. 2009. Packaging Property: The Effects of Paradigmatic Framing of Property Rights. *Tulane Law Review* 83:691–734.

- Nash, J. R., and S. M. Stern. 2010. Property Frames. *Washington University Law Review* 87:449–504.
- Neale, M. A., and M. H. Bazerman. 1991. *Cognition and Rationality in Negotiation*. New York: Free Press.
- Neale, M. A., V. L. Huber, and B. Northcraft. 1987. The Framing of Negotiations: Contextual versus Task Frames. *Organizational Behavior and Human Decision Processes* 39:228–41.
- Novemsky, N., and D. Kahneman. 2005a. How Do Intentions Affect Loss Aversion? *Journal of Marketing Research* 42:139–40.
- Novemsky, N., and D. Kahneman. 2005b. The Boundaries of Loss Aversion. *Journal of Marketing Research* 42:119–28.
- Penner, J. E. 1996. The Bundle of Rights Picture of Property. *UCLA Law Review* 43:711–820.
- Penalver, E. M., and S. K. Katyal. 2007. Property Outlaws. *University of Pennsylvania Law Review* 155:1095–186.
- Plott, C. R., and K. Zeiler. 2005. The Willingness to Pay–Willingness to Accept Gap, the “Endowment Effect,” Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. *American Economic Review* 95:530–45.
- Plott, C. R., and K. Zeiler. 2007. Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory? *American Economic Review* 97:1449–66.
- Polinsky, A. M. 2003. *An Introduction to Law and Economics*. 3rd ed. New York: Aspen.
- Posner, R. A. 2011. *Economic Analysis of Law*. 8th ed. New York: Aspen Publishers.
- Rachlinski, J. J., and F. Jourden. 1998. Remedies and the Psychology of Ownership. *Vanderbilt Law Review* 51: 1541–82.
- Radin, M. J. 1982. Property and Personhood. *Stanford Law Review* 34:957–1016.
- Radin, M. J. 1986. Residential Rent Control. *Philosophy and Public Affairs* 15:350–80.
- Radin, M. J. 1988. The Liberal Conception of Property: Cross Currents in the Jurisprudence of Takings. *Columbia Law Review* 88:1667–96.
- Ritov, I. and J. Baron. 1990. Reluctance to Vaccinate: Omission Bias and Ambiguity. *Journal of Behavioral Decision Making* 3:263–77.
- Ritov, I. and J. Baron. 1992. Status-Quo and Omission Biases. *Journal of Risk and Uncertainty* 5:49–61.
- Rose, C. M. 2000. Left Brain, Right Brain and History in the New Law and Economics of Property. *Oregon Law Review* 79:417–92.
- Roth, A. E. 1995. Bargaining Experiments. In *The Handbook of Experimental Economics*, ed. J. H. Kagel, and A. E. Roth, 253–348. Princeton, NJ: Princeton University Press.
- Samuelson, W., and R. Zeckhauser. 1988. Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty* 1:7–59.
- Schnably, S. J. 1993. A Critique of Radin’s Theory of Property and Personhood. *Stanford Law Review* 45:347–407.
- Serkin, C. 2009. Existing Uses and the Limits of Land Use Regulations. *New York University Law Review* 84:1222–91.
- Shapiro, S., and E. F. McClennen. 1998. Law-and-Economics from a Philosophical Perspective. In *The New Palgrave Dictionary of Economics and the Law*, vol. 2, ed. P. Newman, 460–65. New York: Macmillan Reference.
- Shavell, S. 2004. *Foundations of Economic Analysis of Law*. Cambridge, MA: Harvard University Press.
- Simonson, I., and A. Drolet. 2004. Anchoring Effects on Consumers’ Willingness-to-Pay and Willingness-to-Accept. *Journal of Consumer Research* 31:681–90.

- Singer, J. W. 2000. *Entitlement: The Paradoxes of Property*. New Haven: Yale University Press.
- Singer, J. W. 2005. *Introduction to Property*. 2nd ed. New York: Aspen.
- Smith, H. E. 2012. Property as the Law of Things. *Harvard Law Review* 125:1691–726.
- Stake, J. E. 2001. The Uneasy Case for Adverse Possession. *Georgetown Law Journal* 89:2419–74.
- Stern, S. M. 2006. Encouraging Conservation of Private Lands: A Behavioral Analysis of Financial Incentives. *Arizona Law Review* 48:541–84.
- Stern, S. M. 2009. Residential Protectionism and the Legal Mythology of Home. *Michigan Law Review* 107:1093–144.
- Stern, S. M. 2010. The Inviolate Home: Housing Exceptionalism in the Fourth Amendment. *Cornell Law Review* 95:905–56.
- Stern, S. M. 2011. Reassessing the Citizen Virtues of Homeownership. *Columbia Law Review* 111:890–938.
- Stoebuck, W. B., and D. A. Whitman. 2000. *The Law of Property*. 3rd ed. St. Paul: West.
- Thaler, R. H. 1992. *The Winner's Curse: Paradoxes and Anomalies of Economic Life*. Princeton, NJ: Princeton University Press.
- Thaler, R. J. 2000. Mental Accounting Matters. In *Choices, Values and Frames*, ed. D. Kahneman and A. Tversky, 241–68. New York: Cambridge University Press.
- Thibaut, J., and L. Walker. 1975. *Procedural Justice: A Psychological Analysis*. Hillsdale, NJ: Erlbaum Associates.
- Tian, K., and R. W. Belk. 2005. Extended Self and Possessions in the Workplace. *Journal of Consumer Research* 32:297–310.
- Tor, A., and D. Oliar. 2002. Incentives to Create under a “Lifetime-Plus-Years” Copyright Duration: Lessons from a Behavioral Economic Analysis of Eldred V. Ashcroft. *Loyola of Los Angeles Law Review* 36:437–92.
- Tur-Sinai, O. 2011. The Endowment Effect in IP Transactions: The Case against Debiasing. *Michigan Telecommunications and Technology Law Review* 18:117–69.
- Tyler, T. R. 1989. The Psychology of Procedural Justice: A Test for the Group-Value Model. *Journal of Personality and Social Psychology* 57:830–38.
- Van Dijk, E., and D. van Knippenberg. 1996. Buying and Selling Exchange Goods: Loss Aversion and the Endowment Effect. *Journal of Economic Psychology* 17:517–24.
- Van Dijk, E., and D. van Knippenberg. 1998. Trading Wine: On the Endowment Effect, Loss Aversion, and the Comparability of Consumer Goods. *Journal of Economic Psychology* 19:485–95.
- Weinstein, N. D. 1980. Unrealistic Optimism about Future Life Events. *Journal of Personality and Social Psychology* 39:806–20.
- Weinstein, N. D. 1987. Unrealistic Optimism about Susceptibility to Health Problems: Conclusions from a Community-Wide Sample. *Journal of Behavioral Medicine* 10:481–500.
- Weisbach, D. A. 2003. Should Legal Rules Be Used to Redistribute Income? *University of Chicago Law Review* 70:439–54.
- Zamir, E. 2012. Loss Aversion and the Law. *Vanderbilt Law Review* 65:829–94.
- Zelizer, V. A. 1997. *The Social Meaning of Money*. New York: Basic Books.

## CHAPTER 16

---

# BEHAVIORAL ECONOMICS AND TORT LAW

---

YOED HALBERSBERG AND EHUD GUTTEL

### 1 INTRODUCTION

---

TORT law is the set of legal rules that assigns liability for invasions of legally protected interests, predominantly in cases of accidents, usually with no prior familiarity between the parties. In this chapter we discuss the contributions of cognitive psychology and behavioral studies to the research of tort law. These contributions, we show, relate to a wide range of issues in torts: from the basic decision to impose tort liability, through the choice between liability regimes, to specific tort doctrines and remedies. We also offer some avenues of further research that, in our opinion, need to be explored.

BEHAVIORAL studies are of particular significance for the analysis of the tort system. In retrospect, law and economics largely originated from tort law, with two seminal tort-related papers published in the early 1960s (Coase 1960, Calabresi 1961). Since the 1980s, law and economics and its rational-choice assumption have dominated tort scholarship. This makes torts especially interesting for behavioral analysis and allows for a wide range of applications of insights from the behavioral sciences. Although behavioral analysis of the tort system is only in its nascent stage, it has already influenced both courts and commentators (see, e.g., Restatement (Third) of Torts [American Law Institute 2010, §3 cmt. g]).

The variety and robustness of biases that affect people's assessments of probabilities is another reason for the importance of behavioral economics to tort law. The reason is that probabilities, and the assessments thereof, are key to four facets of tort law. First, accidents are probabilistic in nature, which means both that the parties need to engage in probability assessments before making their preferred choice(s) of action, and that social planners—legislators and courts—need to take the parties' (at times biased) assessments into account when designing and implementing tort

liability.<sup>1</sup> Accordingly, the stochastic nature of accidents has had a direct bearing on the core of the tort doctrine: To determine liability in negligence cases, for example, courts use, either implicitly or explicitly, the Learned Hand formula<sup>2</sup> that requires them to make probability estimates—in a way unparalleled by other areas of the law; in addition, determining liability for design defects in product liability cases requires the court's engaging in estimating risks and benefits that are probabilistic in nature. Consequently, deciding even the simplest tort case exposes the courts—and not just the parties—to a rich array of cognitive biases and heuristics.

Second, tort remedies—especially remedies for bodily injuries—are particularly complex. Punitive damages, damages for pain and suffering, and damages for future harm—all involving the court's having to make rough estimations, probability evaluations of uncertain future events, as well as moral judgments—are prone to biases. Furthermore, awarding damages for pain and suffering requires the court to delve into evaluations of the victims' psyche. Accordingly, the determination of damages in torts may be more susceptible to heuristics and biases than, for example, the determination of punishments in criminal law (for which there usually exist at least upper bounds) or the award of damages in contract law (where the amount of harm is usually less uncertain than in torts).

Third, the parties in tort claims are both more heterogeneous than, for example, parties to corporate litigation, and more unknown *ex ante* than, for example, parties to contract cases. This means that, in torts, parties need to engage in *ex ante* predictions as to the identity of the other side—these predictions being particularly exposed to heuristics and biases; moreover, when the number of victims or injurers is larger than one, their heterogeneity might impinge on the efficiency of assigning tort liability. As a result, biases that affect the perception of group variability may have dire consequences for the efficiency of the assignment of tort liability.

Fourth, attorney fees in tort cases are predominantly based on contingent fees, whereas in other branches of the law fixed fees or hourly fees are more common (Note 1993, p. 450; Zamir and Ritov 2011). This contingent mode of payment makes lawyers' decisions surrounding the litigation and settlement of tort cases, and not just plaintiffs', vulnerable to heuristics and biases in torts, more so than in other areas of the law (because issues of litigation and settlement are analyzed in chapter 24 by Robbennolt in this volume, we will not address them here).

<sup>1</sup> Although probabilities do play an important role in many areas of law (for example, the probability of enforcement in criminal law), in torts they are embedded in the event itself. In torts, probabilities are important even if the enforcement level is at 100%.

<sup>2</sup> The Learned Hand formula for negligence, articulated in *United States v. Carroll Towing Co.*, 159 F.2d 169 (2d Cir. 1947), determines that “failure to take a precaution is negligent. . . if the cost of the precaution. . . is less than the probability of the accident that the precaution would have prevented multiplied by the loss that the accident if it occurred would cause.” (*Mesman v. Crane Pro Servs.*, 512 F.3d 352, 354 (7th Cir. 2008)).



This chapter is organized as follows. Section 2 provides a review of the major contributions of behavioral economics to the research of tort law. It starts with a very brief elaboration of the standard law and economics model of tort law, and then examines the literature's contributions to three key elements: the choice between liability regimes (negligence, strict liability, comparative and contributory negligence); the choice between tort liability and regulation (including the choice between harm- and risk-based liability); and damages (in particular, punitive damages and damages for pain and suffering). In selecting the literature to be included in the review, we have focused on experimental, rather than empirical, research. We do, however, make several references to studies that, although they do not include experimental work, nevertheless relaxed the rationality assumption of the standard economic model and analyzed it under the assumptions of prospect theory and the findings of the behavioral sciences.

Section 3 explores underresearched areas of tort law to which, we think, behavioral economics can make significant contributions. We shall offer two new avenues for future research: vicarious liability, and people's perceptions of the variability among large groups of victims.

## 2 LITERATURE REVIEW: THE APPLICATION OF BEHAVIORAL ECONOMICS TO TORT LAW

---

### 2.1 A Bird's-Eye View of the Standard Economic Model

To fully appreciate how behavioral economics adds to the understanding of the tort system, it might be helpful first to portray, in rough brushstrokes, the basic neoclassical economic model for accidents.

The cornerstone for the basic model was laid first by Calabresi (1961, 1970), and later developed by the seminal works of Brown (1973), Posner (1972a, 1972b), and Shavell (1980). Calabresi clarified that, from an economic perspective, tort law should aspire to minimize three different types of costs: primary costs that are the sum of the expected harm from the accident and the costs of precautions taken by the parties prior to the accident; secondary costs that are incurred in the process of spreading the risk; and tertiary costs of administrating the tort system. The primary economic goal of tort liability is not to compensate victims *per se*—a function best served by first-party insurance; rather, it is to provide the parties with incentives to act in a manner that is aligned with aggregate social welfare, namely, to minimize the social costs of accidents (Posner 1972a, p. 33; Shavell 1980, p. 1). The standard model assumes two agents: an injurer whose activity is risky, and a victim who suffers harm if the risk from the injurer's activity is materialized. The risk to which the victim is subject decreases with the investment in precautions but at a diminishing rate; accordingly, there exists an efficient level of



precautions that minimizes the overall social costs of the accident, such that taking either more or less care is socially undesirable.

In analyzing tort liability, it is customary to distinguish between three different categories of accidents according to the precaution-technology available to the parties: *unilateral-care situations*, where only the injurer can effectively take precautions that lower the expected harm; *alternative-care situations*, where the precautions that are available to the victim are substitutes to those of the injurer; and *bilateral-care situations*, in which the victim's precautions are supplementary to the injurer's, so that both are needed to efficiently reduce the expected harm. From an economic perspective, in unilateral-care cases, tort liability should aspire to make the injurer internalize the harm, thereby driving her to minimize the sum of the expected harm and the costs of precautions. In alternative-care cases, tort law should be construed to induce only the accident's least-cost avoider to invest in precautions (Shavell 1987, p. 18; Posner 2007, p. 40). In the bilateral-care case, the least-cost avoiders are either both parties together or none of them.

It is also customary to distinguish between levels of care taken by the parties, that is, the amount (or cost) of precautions in which the parties invested, and levels of activity that, traditionally, mean either the number of times that parties engage in the risky activity, or investment in precautions that are invisible to the court (Shavell 1980; Posner 2007, p. 178). Thus, in driving a car, the quality of the brakes and the driving speed, for example, belong to the level-of-care category, whereas the number of miles driven is part of the level of activity. Similarly, in industrial manufacturing, the amount and the quality of a factory's chimney filters is part of the factory's level of care, while the number of hours the factory operates is its level of activity.

The *standard* model is based on a series of assumptions, some of which were later relaxed in economic (rational-choice) models, and others by behavioral research. The basic assumptions are that the agents are rational and risk-neutral, that they (and the courts) have perfect information, that transaction costs are prohibitively high so that parties cannot negotiate prior to the tortious activity; that courts do not make systematic errors; and that the damage awards exactly equal the amount of harm. It is also generally assumed that, when determining negligence, courts see the case through the lens of aggregate social welfare; hence, they set the standard of care at the socially efficient level.

## 2.2 Tort Liability Regimes

### 2.2.1 *The Basic Insights of the Standard Economic Model*

The standard model's starting point is the absence of tort liability. Under such a no-liability regime—if transaction costs are prohibitively high (Coase 1960)—the injurer will take no precautions, and engage in too high a level of activity. The reason for this is that the injurer suffers no harm as a result of her own activity and precautions

are costly; accordingly, the injurer's goal of minimizing her private costs is achieved by minimizing her costs of care, that is, by taking no precautions. On the other hand, in a no-liability world, given that the injurer does not take precautions, the victim will take her socially efficient level of care—since the victim suffers all the harm without compensation, and thus bears the social costs of the accident in their entirety: both the cost of precautions and the expected harm. In alternative-care situations where the victim is the accident's least-cost avoider, this induces the first-best outcome, but in bilateral-care cases, unilateral-care cases, or when the injurer, rather than the victim, is the least-cost-avoider—this results only in a second-best outcome.

The opposite regime to that of no liability is the rule of strict liability. Under this rule, the injurer compensates the victim whenever the risk materializes and harm is inflicted upon the victim—with no regard to fault. Under strict liability, in a unilateral-care setting, the injurer takes efficient precautions and engages in the efficient level of activity; the reason is that, after compensating the victim, the injurer bears all social costs. In bilateral- or alternative-care settings, however, the victim does not take any precautions and may engage in too much activity. This is due to the victim's bearing only the cost of her precautions, which may not affect the injurer's liability toward her. In sum, the rules of no liability and strict liability are similar in that they can efficiently incentivize only one party: either the victim or the injurer, but not both. Consequently, in unilateral- and alternative-care situations these regimes might fail to achieve optimal deterrence—depending on whether the injurer or the victim is the accident's efficient harm-avoider—whereas in a bilateral-care setting they necessarily fail (Shavell 2004, Schäfer and Schönenberger 2000).

In modern tort systems, however, examples of a no-liability regime and of pure strict liability (with no defenses) are rare. Most tort-related activities are subject to some form of a negligence-based rule, be it pure negligence, comparative negligence, strict liability with the defense of contributory negligence, or negligence with the defense of contributory negligence. The core of these negligence-based rules is a risk-benefit test, namely, the Learned Hand formula that weighs the ex ante marginal expected harm against the marginal cost of precautions. Of relevance to our discussion is that the “expected harm” component of the Hand formula embeds probability assessments. A crucial assumption in the standard model is that the fact-finder makes accurate probability assessments and accurate harm assessments, so that the court determines the standard of care correctly. Anchored in this assumption is the standard model's result that all negligence-based regimes drive both the injurer and the victim to take the socially efficient level of care in unilateral-care, alternative-care, and bilateral-care cases alike. The reason for this is that when a party takes the socially optimal level of care, it is deemed nonnegligent; thus, it is in the interest of the other party, too, to take the efficient level of care. On the other hand, with regard to the parties' *levels of activity*, the negligence-based rules yield nonidentical results: Only the party that bears the residual harm (i.e., the victim under negligence and the injurer under strict liability) can be driven to engage in the efficient level of activity (Shavell 1987, pp. 41–46). The reason is that the parties take their efficient level of care, and therefore are not held negligent or

contributorily negligent. Accordingly, under a negligence regime the risk created by the injurer's excessive activity is fully externalized to the victim, and vice versa under strict liability.<sup>3</sup>

Some assumptions of the basic model were relaxed in later economic models. For a review of such relaxations see, for example, Cooter and Ulen (2012, ch. 7) who discuss the implications of insurance, injurer's solvency, litigation costs, courts' random errors, and evidentiary uncertainty. They also succinctly discuss relaxing the rationality assumption, noting that injurers or victims may be biased in their probability assessments of the risk, thereby making the tort system incapable of inducing efficient behavior. For a further review see Schaefer and Mueller-Langer (2008).

### 2.2.2 *Merging the Insights of Behavioral Economics into the Standard Model*

The first to recognize that the choice between the different liability rules is affected, inter alia, by human psychology were Calabresi and Hirschoff (1972, p. 59) who noted that the relative effectiveness of liability regimes depends on the parties' de facto ability to process information and to minimize accidents' costs; this, in turn, depends (among other factors) on "the absence of psychological or other impediments to acting on the basis of available information." Other law and economics scholars, too, have emphasized that the standard economic model assumes rationality and, therefore, if parties are not rational the conclusions from the standard model might change (Shavell 1987, pp. 74–82; Cooter and Ulen 2012, ch. 7; See also Faure 2010, pp. 20–21 for review).

In addition to *the parties'* psychology, the choice between the different liability rules should depend on *the courts'* own cognitive biases. Determining negligence, or contributory negligence, requires the court to decide on whether the costs of taking precautions were lower than the risk that those precautions could have averted. Assessment of risk, however, entails probability evaluations that are prone to biases. Strict liability, on the other hand, does not necessitate probability evaluations; hence, the hazard of cognitive biases leading to judicial errors is greater under negligence than under strict liability (Korobkin and Ulen 2000), and, similarly, greater under negligence with the defense of contributory negligence than under strict liability with the same defense.

Two related biases in the estimation of probabilities and assignment of fault, which may affect the choice between liability rules, are the hindsight bias and the outcome bias. The hindsight bias, discovered by Fischhoff (1975), distorts people's ex post assessments of the ex ante probability and predictability of an event, given that this event has already happened (chapter 14 by Teichman in this volume). The outcome bias is the tendency to perceive conduct that resulted in a bad outcome as more careless than the same conduct in cases where the bad outcome did not occur. Despite their similarity in distorting ex post evaluations of behavior, the hindsight and outcome biases are

<sup>3</sup> Gilo and Guttel (2009) show, however, that increased risk may also stem from insufficient, rather than excessive, activity.

distinct in that the outcome bias does not relate to probabilities as such, but rather to the evaluation of the quality and reasonableness of decisions; it may therefore appear even with certain knowledge of the ex ante probabilities (Peters 1999, p. 36). The hindsight bias's ramifications for the law of torts were explored and explained in Kamin and Rachlinski (1995) and Rachlinski (1998). Kamin and Rachlinski (1995) presented undergraduate students with a case similar to the facts of *Petition of Kinsman Transit Co.*, (338 F.2d 708, 2d Cir. 1964). The case surrounded a drawbridge and a city's determination of whether the risk of flood called for a bridge operator to be hired in winter when the bridge was not in use. Participants in the experiment were divided into three groups: a hindsight group who were told that a flood had indeed occurred and that maintaining a bridge operator would have prevented harm to a nearby bakery; a control group who were asked to decide in foresight whether to maintain a bridge operator; and a debiasing group who were also given instructions to disregard the fact that the flood had happened. While only 24% of the foresight group chose to hire an operator, 56% of the hindsight group and 57% of the debiasing group decided that an operator should have been hired. These results demonstrate both that the hindsight bias is present in the evaluations of tort cases and that simple mechanisms such as jury instructions are insufficient as debiasing tools.

The tort-related outcome bias was demonstrated in Baron and Hershey (1988). The authors told participants that a heart bypass operation entails a risk of 8% of dying, and asked them to determine whether a decision to operate was reasonable. Participants who had been exposed to data about the operation failing viewed the surgeon's decision to operate as much less reasonable than participants who were told that the operation was successful.

The above experiments explored either hindsight bias or outcome bias. The impacts of both these biases on tort-related cases were demonstrated in Labine and Labine (1996). The authors sent questionnaires to registered voters (eligible to serve as juries), asking them to examine the ex ante foreseeability of a mental health patient turning violent, and to determine whether the mental health professionals who treated that patient acted unreasonably (and, therefore, would be held negligent if the case were tried as a negligence case). Among the 297 respondents, those who were told that the patient had indeed become violent perceived the violence as more foreseeable, and the professionals more negligent, than respondents who were told either that the patient had not become violent or were not told of the patient's behavior.

The outcome and hindsight biases have a direct bearing on the choice between liability rules. The negligence doctrine, and in particular the Hand formula for negligence, require the judicial fact-finder to do exactly what the hindsight bias affects: to make ex post assessments of the ex ante probability of the accident occurring. Naturally, then, even doctrinal legal scholarship sometimes referred to the behavioral scholarship on the hindsight bias (American Law Institute 2010, §3 cmt. g). Furthermore, the courts determine injurers' liability for an accident only after it has occurred, leaving large room for the outcome bias to take effect. In contrast, strict liability does not necessitate estimations of probability and of foreseeability of events, nor does it require the court to

evaluate the carelessness of an actor whose conduct resulted in harm; hence, strict liability may lead to less biased results. Consequently, when weighing strict liability against negligence, the hindsight bias may tilt the balance in favor of strict liability (Korobkin and Ulen 2000).

It should be emphasized, however, that tort cases under strict liability are not entirely immune to the impact of the hindsight bias and the outcome bias, the reason being that an accident's foreseeability and, accordingly, the injurer's carelessness are factors that juries weigh when determining punitive damages. This effect was demonstrated in Hastie, Schkade, and Payne (1999b). The authors conducted two experiments in which subjects were presented with the facts of a tort case involving an environmental accident of a train derailing and dumping toxic herbicide into a river. Some subjects were told that the accident occurred and some were asked to foresee the chances of the accident happening; also, some subjects were assigned juror roles. The results showed that the accident was perceived as more foreseeable when the subjects were told that an accident had indeed occurred, and, accordingly, the railroad was judged more severely and was liable for more punitive damages under this treatment.<sup>4</sup> Since strict liability is more sensitive than negligence to errors in overestimating damages (Posner 2007, p. 181), and since the hindsight and outcome biases affect the award of punitive damages, there is also a case to be argued that the overall impact of the two biases in favoring strict liability over negligence is unclear, and more research on this point is still needed.

Building on the above experimental results, a burgeoning theoretical literature has examined the implications of the hindsight bias on the information jurors should be provided in tort-related cases. Most notably is Jolls, Sunstein, and Thaler (1998) who propose—in order to prevent the overestimation of the accident's probability caused by the hindsight bias—to withhold information from juries as to what actually happened and to raise the burden of proof for ascertaining negligence. In contrast, Rachlinski (1998) and Peters (1999) emphasized that the implications for tort law of the hindsight bias are more limited than one can expect, since current law already provides debiasing mechanisms, such as the Subsequent Remedial Measures Rule<sup>5</sup> that excludes evidence of subsequent remedial measures to prove the defendant's negligence, arguments by the defendant's counsel, and prodefendant biases (such as jurors' distrust of personal injury plaintiffs and obstacles that prevent victims of negligence from filing successful claims).

Theoretical scholars embedded other biases, too, into the standard economic model, and, using these enhanced models, came to different conclusions. Posner (2004)

<sup>4</sup> The participants did, however, demonstrate smaller hindsight effects as jurors than as simple citizens. This corresponds both with Viscusi (1999), who demonstrated, using written questionnaires, that judges exhibited smaller hindsight effects than the general public, and with Hastie and Viscusi (1998), who also showed that in tort cases judges may be less prone to the hindsight bias than jurors. The research in Hastie and Viscusi (1998) gave rise to a methodological debate: Lempert (1999) criticized their methods, and the authors replied in Hastie and Viscusi (2002).

<sup>5</sup> Fed. R. Evid. 407.

incorporates both the overoptimism bias and probability insensitivity (i.e., the insensitivity to small differences in probabilities) into the standard unilateral-care model, and reaches some unintuitive conclusions. His study shows, first, that these probability errors have similar ramifications for the two main liability regimes (negligence and strict liability with the defense of contributory negligence). Next, Posner divides overoptimism into three categories: high levels of optimism, low levels of optimism, and moderate levels of optimism. He shows that low levels of optimism do not alter the results of the standard model that is based on the assumption of rationality, in that the two regimes induce efficient care by the parties. Under high levels of optimism, however, the injurer may take less than optimal care, and engage in too much activity. Surprisingly, in moderate levels of optimism the injurer may take too *much*, rather than too *little*, care. The reason is that an optimistic injurer may overestimate the efficacy of precautions, and thus may think that taking little extra care will have dramatic effects on her expected liability. Consequently, she may be driven to take extra care that a rational injurer would not perceive as beneficial. As for bilateral care situations, Posner concludes that with optimism, both injurers and victims will take too little care, but also that the difference between the two liability rules with regard to levels of activity—which is one of the main results of the standard economic model—vanishes with optimism: the two liability rules drive both the injurer and the victim to engage in too much activity.

Theoretical scholars have also explored liability rules within the framework of prospect theory. Specifically, Bigus (2006) applies prospect theory's probability-weighting function to the standard model of unilateral-care situations, assuming that the injurer overestimates very low probabilities and underestimates very high ones (Kahneman and Tversky 1979). The result of the weighting function is that the injurer's marginal utility from taking care decreases, which leads to underinvestment in precautions. This underdeterrence effect occurs under strict liability, whereas under negligence its existence depends on the slope of the probability-weighting function. Bigus also finds that when the standard of due care under negligence is vague, probability weighting drives the injurer to take more care, thereby alleviating the underdeterrence effect of ambiguity and even inducing too much care.

Teitelbaum (2007) adds to these models by distinguishing between cases of fixed accident losses (where precautions affect only the probability of harm) and variable accident losses (where precautions also affect the magnitude of the harm). Using the Choquet Expected Utility Theory (see Sugden 2002, p. 734, for review) to model risk ambiguity, Teitelbaum finds that neither strict liability nor negligence is generally efficient, but negligence is more ambiguity-proof. In the fixed-losses scenario, injurers may take too little care under negligence, while under strict liability injurers *always* take too little care. In the variable-losses scenario, negligence may drive the injurer to take too little care, while under strict liability she may take either too much or too little care. Teitelbaum also shows that the injurer's level of care generally decreases with optimism and increases with pessimism, while the effects of ambiguity itself on the injurer's level of care depend on whether the injurer is optimistic or pessimistic.



Bhole and Wagner (2010) add to Posner's (2004) and Teitelbaum's (2007) models by showing that punitive damages may be a means for achieving optimal deterrence when agents are optimistic or when the standard of due care used by the courts is ambiguous.

While the hindsight and outcome biases may tilt the scales in favor of strict liability, these theoretical models seem to favor the negligence rule over strict liability. The theoretical models, however, should be considered with caution. The difficulty is that they naturally focus on one or two aspects of one or two cognitive biases, whereas the richness of behavioral phenomena may substantially limit the models' external validity, similarly to experiments that isolate only one bias. For example, Posner's model (2004), although insightful, deliberately focuses on a special form of overoptimism: overoptimism with regard to the effectiveness of precautions taken by the injurer. The conclusion that moderate levels of optimism can induce more, rather than less, care, should therefore be taken in context. In real life, the optimism with regard to precautions' effectiveness may interact in the injurer's mind with the "it will never happen to me" kind of overoptimism, and the cumulative effect of these two facets of overoptimism may be less obvious. Indeed, Posner concludes that when exploring overoptimism, a specific elaboration on what this overoptimism relates to (the probability of an accident, efficacy of precautions, and so forth) is required. This, of course, takes only overoptimism into account and disregards other behavioral phenomena. The availability heuristic, for example, might increase awareness about certain categories of accidents but not others—which may affect the efficiency of precautions that lower the risk of one type of accident but increase the risk of other, less cognitively available, events.

It is, therefore, especially interesting to review the results of experiments that can capture the cumulative effects of a whole range of behavioral biases and heuristics on the participants' minds under different liability rules settings—similarly to real-life agents. In two studies conducted in the early 1990s, Kornhauser and Schotter (1990; 1991) performed laboratory experiments that explored how different liability rules affect participants' decision-making. In the earlier paper, their multiround laboratory experiment explored the incentives that negligence and strict liability provide to injurers in a unilateral-care setting. Surprisingly, the authors found no support for the predictions of the standard model as to the equivalence between the two rules. Rather, they found that in the earlier rounds participants took more care under strict liability than under negligence, while in later rounds the subjects took less care under strict liability than under negligence. Furthermore, Kornhauser and Schotter found that, contrary to what the standard model predicts, participants did not choose a significantly higher level of activity under negligence than under strict liability. The comparison between the two rules indicated that negligence consistently outperformed strict liability, and was more efficient both when the standard of due care was set at the efficient level and when it was set at too high a level. The authors hypothesize that under strict liability participants found it hard to compute the efficient level of care, while under negligence the standard of care helped participants find the efficient level.



In the 1991 paper, Kornhauser and Schotter explored the bilateral-care setting, comparing pure negligence on the one hand and negligence with the defense of contributory negligence on the other. The authors found that when the standard of due care is set optimally, the two liability rules perform more or less equally well, and that under pure negligence participants who bear liability in equilibrium<sup>6</sup> tend to take excessive care at the early rounds and less-than-optimal care at later rounds. In this respect, negligence with the defense of contributory negligence seemed preferable to pure negligence. The authors did, however, find some support for several theoretical predictions: Although both regimes were not effective in inducing the parties to engage in their efficient level of activity, parties who could be liable for the harm did engage in less activity than when they were not subject to the threat of liability. Moreover, the experiment demonstrated that under both rules, participants had reacted very sensitively and efficiently to changes in the standard of due care. Surprisingly, however, participants' knowledge about the level of care taken by the opposing party did not improve the effectiveness of either rule to induce optimal deterrence.

Another study that tested the efficiency of different liability rules is Ghosh and Kundu (2010), who compared strict liability to no liability. This study found that under strict liability "injurers" took more care than is efficient. The authors also found, surprisingly, that under a no-liability rule the "injurers" did take some care—their lack of liability notwithstanding. Before drawing conclusions from their results, however, it should be noted that each treatment lasted only two rounds. This may be a source of skepticism because, as described earlier, in Kornhauser and Schotter's studies (1990; 1991) the participants changed their behavior over time. Nevertheless, Ghosh and Kundu's surprising result that injurers take precautions even in a no-liability world receives some support from both dictator games<sup>7</sup> and another study, conducted by Angelova, Attanasi, and Hiriart (2012). In the latter experiment, the researchers explored the effects of strict liability, negligence, and no liability, in unilateral-care situations, on firms whose activity poses a risk of catastrophic harm on the environment or a large number of victims (who are neither consumers nor employees). The researchers were especially interested in the role that insolvency plays in the firm's decision-making process. The theoretical, neoclassical prediction is that when an injurer firm is subject to risk of insolvency if it is liable to pay damages, the negligence rule performs better in inducing it to take efficient care<sup>8</sup> (Landes and Posner 1987, pp. 258–72) and

<sup>6</sup> Namely, the injurer when in equilibrium the injurer takes less care than the standard of care, and the victim when in equilibrium the injurer takes more care than the standard.

<sup>7</sup> A no-liability rule is similar to a dictator game, because the injurer can one-sidedly decide whether she wants to benefit the victims by taking precautions. And, just as in dictator games, where the behavioral phenomenon is that dictators usually give some of the money to the subject (chapter 2 in this volume by Simon Gächter), in a no-liability world it seems that injurers do take some precautions.

<sup>8</sup> Under negligence, by taking efficient care the injurer avoids liability completely, and thus it may avoid bankruptcy by taking efficient care, while under strict liability the injurer might still pay large amounts as damages even if it took efficient care.

the injurer's level of care is lower than under no risk of insolvency. Surprisingly, however, the researchers found no support for insolvency playing any significant role in the participants' decision-making processes. Furthermore, they found that although investment in precautions under strict liability and negligence was always higher than under a no-liability rule, even under a no-liability rule the injurers did tend to take some precautions. Moreover, contrary to Kornhauser and Schotter (1990), they found that investment in precautions is not significantly different under negligence than under strict liability, although under negligence the participants tended to take *insignificantly* more care. We think that this result might be explained by the fact that in Angelova, Attanasi, and Hiriart's experiment, the participants faced a fairly obvious choice: It was quite clear to the participants that taking precautions is efficient and that the injurer benefits from taking precautions under the strict liability rule. Nonetheless, both Kornhauser and Schotter's and Angelova, Attanasi, and Hiriart's experiments provide some support for what might be described as the positive effects of a "semianchoring" effect made by the standard of due care under the negligence rule.

In Angelova, Attanasi, and Hiriart (2012), the pattern of behavior under strict liability, over time, repeated Kornhauser and Schotter (1990): Participants started by overinvesting in precautions in the early rounds, but reduced their level of care in later rounds. This learning curve and convergence to equilibrium were the focal point of Wittman and his coauthors' study (1997). Here, the authors explored a setting of symmetrical harm (namely, when the accident occurs, both parties suffer the same amount of harm), and compared the subjects' chosen levels of care under comparative negligence, negligence with the defense of contributory negligence, and no liability. The authors explain that due to the symmetry in damages, the no-liability rule is equivalent to strict liability. (We doubt this last assumption, as framing effects and moral intuitions may cause agents to perceive no liability and strict liability differently even with symmetrical harm. This specific point warrants further research.)

Wittman and his coauthors' main interest was the question which liability rule induces a behavior that converges quicker to equilibrium. This question is of great importance, because the shorter the route to an efficient equilibrium, the less overall social cost is incurred. The experimental results demonstrated that, as predicted by the neoclassical model, the no-liability rule induced the participants' taking less-than-optimal care (yet, more than the standard-model prediction of no care whatsoever), while comparative negligence and negligence with the defense of contributory negligence induced relatively close-to-optimal behavior. However, contrary to the classical model, behavior under contributory negligence converged more closely and more quickly to equilibrium than under negligence with the defense of contributory negligence. The authors therefore provided a theoretical explanation, supported by experimental evidence, for the prevalent use of comparative negligence rules in modern tort systems: Comparative negligence induces efficient behavior faster and more accurately than other liability rules.

As with most laboratory experiments that try to mimic real-life situations and draw legal conclusions, the above-described experiments are exposed to criticism regarding

their external validity. In particular, there are three main points of difference between real life and tort-liability laboratory experiments. First, the participants in experiments are mainly young students. Second, although there are monetary ramifications for subjects' decisions, the participants do not have much skin in the game, and, contrary to tort situations, participants may not suffer *real loss* in the experiment, but rather, at the most, loss of possible profit. Third, these experiments did not explore the tort system in its entirety, with biases that affect the injurers, the victims, and the courts. It is surprising, that in spite of the extensive attention that the neoclassical literature has devoted to the difference between the various liability rules, there has been little experimental work that explores how these rules work in reality. Further research is, therefore, still needed to enhance our understanding of the differences between the major liability rules and to build a comprehensive behavioral theory of liability rules.

### 2.3 Regulation versus Tort Liability: Standard Economic Considerations for Choosing between Regulation and Tort Liability

Inducing efficient behavior by injurers, by making them internalize the risks they impose on others, can be achieved either by a threat of tort liability or by administrative means, such as direct regulation of activity (followed by criminal or administrative sanctions) or imposition of Pigovian taxes. The choice between regulation-based incentives and tort-based incentives has sparked both economic and political debates. The economic perspective was laid down mainly by Wittman (1977) and Shavell (1984, pp. 358–64), who reviewed the often-conflicting considerations surrounding the choice between the two:

- a. Information gaps between the regulator and the parties and courts, as to the costs and benefits of the regulated activity: The more the parties know than the regulator, the better tort liability is than regulation, and vice versa.
- b. The potential injurers' solvency capacity, and the existence of insurance: the greater the risk of injurers' insolvency, the more society should prefer regulation to tort liability, the reason being that under a regulatory regime, the injurers are subject to *ex ante* liability, and are required to obey the standard or pay a fine (or a tax) before harm materializes. Thus, the size of the fine or tax is usually smaller than the harm, as it is discounted by the probability of the harm materializing.
- c. The chance of a successful lawsuit in case the harm occurs. The smaller the scale of private enforcement the better a regulatory solution is.
- d. The relative costs of administrating the tort system as opposed to the regulatory system. When the costs of data collection are not too high, regulation may be relatively cheap, since it will not require long and complex litigations. In contrast,

tort liability is only imposed if the risk materializes, and—under negligence—if the injurer is at fault. As a result, the costs of enforcement may be higher under a regulatory regime, where enforcement occurs in 100% of the cases (Shavell 2013).

### 2.3.1 *Why Behavioral Economics Matters*

The findings of behavioral economics have a direct bearing on the choice between the alternatives of tort liability and regulations. They affect the decision both by influencing the above considerations and by introducing other, new, considerations.

Applying the behavioral findings to the first consideration above (the informational gaps between the regulator and the parties) may suggest that regulation is superior to tort liability. This is not because regulators themselves, as people, are less vulnerable to biases, but rather because the regulation process is more structured, and usually made based on technical tests that may be less prone to biases than parties' decisions. Adler and Posner (1999) and Sunstein (2000) assert that cost-benefit analyses—which are at the core of the regulation process—can serve as a policy-fixing tool and a counterpower that balances erroneous perceptions that stem from cognitive biases.

In addition, regulation can induce better deterrence due to its clear standards. As explained above, Kornhauser and Schotter's (1990; 1991) results suggest that negligence may be better than strict liability in that its standard of care helps parties explore their best strategy of behavior. Similarly, regulation may be superior, in this respect, to negligence—since the rules that regulation creates may be even less ambiguous than standards of care.

Furthermore, framing effects may also affect the relative deterrent effects of regulations and tort liability. As Feldman and Teichman (2008) showed, fines induce a perception of the sanction as a price—and this, in turn, may encourage agents to indulge in private risk-benefit analyses. If efficiency is our primary role, regulations that impose fines may be better than tort damages in achieving optimal deterrence. However, in cases where an activity is altogether undesirable, fines may be less effective.

The other considerations—risk of insolvency, rates of insurance, and probability of successful lawsuits if harm occurs—seem more ambiguous. First, it is necessary to understand that what actually plays a role here is not the real probability of insolvency but rather the perception of this risk in the minds of potential parties, both injurers and victims. As for victims' biases, on the one hand, the availability heuristic may make victims overestimate the risk of injurers' insolvency, leading to overinvestment in victims' precautions in alternative-care situations when investment in the injurer's precautions is more socially desirable. On the other hand, victims' overoptimism might push the other way. Potential injurers' perception of the risk of insolvency may also be distorted by cognitive biases, but the precise direction is unclear. Overoptimism may cause injurers to discount the risk of insolvency (Harris and Albin 2006, regarding private injurers; Crane 2011, regarding managers of firms)—which may in itself lead to more

deterrence generated by tort liability<sup>9</sup>—but at the same time may also induce underestimation of both the risk of harm and of tort claims filed against them—thereby leading to less than optimal deterrence. In this respect, it is interesting to refer once more to Angelova, Attanasi, and Hiriart (2012), who found no effect for risk of insolvency on the “injuror’s” level of care. Furthermore, the impact of the availability heuristic makes these considerations even more ambiguous, as managers of firms might overestimate the risk of insolvency or of harm based on the occurrence of such events to other firms.

But behavioral findings also add other considerations that may argue in favor of tort liability over regulations. These are considerations relating to the political economy of regulation. Since regulations are sometimes the result of public demand, the overall demand for risk regulations might be decreased by the overoptimism bias. In addition, the availability heuristic may increase the demand for regulation of specific risks that are more robust in the general public’s mind, and decrease demand for regulating other risks (Noll and Krier 1990; but see Camerer 1990). Furthermore, consumers of dangerous products, for example, may underestimate the benefit from safety regulations (Viscusi 1999). The above may, in turn, distort the political incentives provided to regulators and make them suboptimal. Indeed, analysis of empirical data suggests that regulations are affected by political variables and risk-perception biases, and this might make regulations inefficient (Hamilton and Viscusi 1999). Political-economy considerations may also qualify arguments as to the relevance of the hindsight bias to the question in point. Scholars have argued that this bias may tilt the balance toward regulation, as regulations are made in foresight—in contrast to tort liability that is imposed *ex post* (Rachlinski 1999; Peters 1999, p. 1292). However, the regulatory process itself may also be subject to the hindsight bias: Political forces may push regulators to “do something” to prevent bad events from reoccurring, even if, in reality, no regulation is required.

### 2.3.2 *Risk- versus Harm-Based Regimes*

One of the main differences between tort liability and regulation is that the former is traditionally harm-based, whereas the latter is risk-based. As opposed to regulation, only when the risk that is entailed in the tortious behavior is materialized into some harm does tort law come into effect.<sup>10</sup> However, from a theoretical point of view, one can imagine a risk-based tort system, where injurers pay into a general or governmental fund whenever their tortious behavior entails some risk of harm; whenever harm materializes, victims receive compensation directly from that fund (Molot 1997; Porat and Stein 2001). Such a risk-based tort system would be less affected by behavioral biases. From the optimal deterrence perspective, both harm- and risk-based systems are identical insofar as probability assessments are accurate (Issacharoff 2002, p. 1076),

<sup>9</sup> This may be another example of the overoptimism bias as deterrence-enhancing, which may correspond with Posner’s (2004) results.

<sup>10</sup> However, in some limited-scope issues, modern tort systems have also applied some risk-based liability, using doctrines such as loss of chance or increased risk of harm. For a review see Guttel and Harel (2005, pp. 1232–33).

but biases in the perception of probability may make a risk-based system preferable to the current harm-based system. The above discussion with regard to the choice between regulation and tort liability is generally applicable also to the choice between harm- and risk-based liability systems (victims' decisions, however, are not affected by the choice between the two liability systems). In addition, Guttel and Harel (2005) have asserted that, since risk-based systems entail payments for risky behavior while enforcement under harm-based regimes depends on actual materialization of harm, the phenomenon known as probability-matching<sup>11</sup> may generate underdeterrence in harm-based systems when injurers engage repeatedly in risky activities.

In sum, more experimental research that compares between regulation and tort liability, and between risk-based and harm-based tort systems is still needed.

## 2.4 Damages in Torts

### 2.4.1 *Pain and Suffering*

Tort damages compensate for several types of losses. Because lost wages and medical expenses—two of the central categories of recoverable losses in torts—largely depend on objective factors (victims' salary and the costs of medical treatment, respectively), they are generally less vulnerable to psychological biases and judgment-related cognitive errors. In contrast, pain and suffering—the third major type of recoverable losses—mostly hinges on jurors' and judges' discretion. Two important behavioral tendencies have been shown to affect this discretion. First, framing effects influence individuals' assessments regarding the proper amount of compensation for pain and suffering. Second, fact-finders' failure to recognize victims' "hedonic adaptation" may induce excessive compensation to victims at the expense of injurers.

### 2.4.2 *Framing*

Pain and suffering damages are intended to compensate victims for the nonpecuniary loss they suffer following their injury. Accordingly, it seems that to determine pain and suffering it is merely required to instruct jurors to translate victims' agony to a dollar amount. As researchers have shown, however, the legal system may apply two alternative approaches that differ in the baseline used to assess victims' harm. Under the first approach, jurors are instructed to gauge how much victims would need to receive to "be made whole." Such an approach takes victims' injury as its reference point and asks ex post what amount will compensate them for their anguish. Alternatively, the legal system may instruct jurors to gauge how much one would demand to sustain pain and

<sup>11</sup> "Probability matching can be defined as the tendency to adopt a mixed strategy dictated by the relative frequency of events, even when the utility-maximizing strategy would be to always behave in a way that presupposes that the most probable event would occur" (Guttel and Harel 2005, p. 1201).

suffering of the kind experienced by the victims. Under this second approach, the baseline is set at the preharm stage and requires jurors to apply an *ex ante* test. While rational choice may ascribe little weight to this difference, it has been shown to significantly affect jurors' determination.

In an influential experiment, McCaffery, Kahneman, and Spitzer (1995) requested law students and laypersons to determine pain and suffering damages in several hypothetical tort scenarios. Participants were divided into three major groups: no instructions, instructions applying the *ex post* approach, and instructions applying the *ex ante* approach. Interestingly, the results showed that the control group (no instructions) awarded the lowest compensation, suggesting that any instruction that makes jurors put themselves in the victims' shoes is likely to increase the compensation. In addition, the results showed that the *ex post* group awarded nearly half the amount awarded by the *ex ante* group. These results were found consistently across the various scenarios and independent of the identity of the participants (students or laypersons). While the experiment results show that the type of instructions matters a great deal, participants themselves were found to be rather oblivious to this fact. In another part of the experiment, participants were presented with the two types of instructions and asked which is likely in their opinion to induce greater compensation to victims. A majority of participants predicted, in contrast to the experiment actual results, that the *ex post* approach will induce greater compensation.

From a theoretical perspective, these results may be directly connected to prospect theory: Considering pain and suffering damages from an *ex ante* perspective frames the question as being about losses, whereas consideration *ex post* frames the issue as being about gains. Since losses loom larger than gains (chapter 11 by Zamir in this volume), consideration *ex ante* results in larger damage awards than consideration *ex ante*.

From a practical perspective, these results suggest that sophisticated plaintiff lawyers can influence their clients' damage awards for pain and suffering by inducing jurors to consider the harm from an *ex ante* rather than *ex post* approach. As McCaffery and his colleagues suggest, the desire to limit this lawyers' influence may explain the restrictions imposed on plaintiffs lawyers in arguing for pain and suffering. Under the so-called Golden Rule, plaintiffs' lawyers cannot request jurors to consider how much they would demand "to be in the victim's shoes." The "Golden Rule" ensures that all jurors, irrespective of their lawyers' cognitive sensitivity, will be subject to the same compensation standard.

### 2.4.3 *Hedonic Adaptation*

Whereas pain and suffering look to compensate victims for the sensation of discomfort that follows a physical injury, and for the various types of emotional upset that results from tortious activities, hedonic damages expand victims' compensation for what is often described as "the loss of enjoyment of life" (see, e.g., *Knight v. Lord*, 648 N.E.2d 617, 623 (Ill. App. Ct. 1995)). An increasing number of jurisdictions now recognize the right of victims, who due to their disability can no longer enjoy "pleasure from normal



activities of daily life,” to receive compensations beyond traditional pain and suffering damages (Poser, Bornstein, and McGorty 2003).

Compensation for hedonic damages is grounded in the recognition that disability deeply affects one’s life. As courts have often emphasized, disability is not limited to its plain physical aspect but rather goes to the very root of conducting a meaningful life. Under this view, the loss of bodily integrity deprives victims of such things as the ability to engage in social activities, to take part in important family events, to exploit their talents and skills, and perhaps most importantly, to maintain a perception of being self-governed individuals. A disabled person thus cannot fully exhaust the opportunities of life and is bound to experience prolonged frustration. Hedonic damages are designed to account for this loss.

The notion that disability is a tragedy in the sense just described has been challenged, however, by psychologists. A robust body of studies shows that victims who experience physical disability tend to adapt to their condition, and as part of their adaptation learn to readjust their life to their postinjury abilities (see. e.g., Frederick and Loewenstein 1999). This adaptation enables victims to minimize the implications of their injury on their subjective well-being. Studies show that victims’ adaptation might even start as early as several weeks after their injury. For example, although people with spinal cord injuries are “very unhappy immediately following their trauma,” most of them report that they are happy by the third week after the accident (Diener and Diener 1996). A long line of studies show that even for severe injuries, victims’ satisfaction with their life rises rapidly and often equals (or nearly equals) that of fully abled individuals (e.g., Stensman 1985).

Consistent with this discrepancy between the courts’ (pessimistic) view and the actual adaptation of disabled individuals, multiple studies have shown that people are poor predictors of how unfamiliar circumstances will affect individuals’ well-being. In what has become probably the most well-known work in this area, Brickman, Coates, and Janoff-Bulman (1978) interviewed lottery winners, accident victims, and ordinary people (a control group) to evaluate the degree to which major positive or negative events affect individual’s happiness. When asked to gauge the effect of each of these events (a lottery winning or a severe accident), people predicted that it would dramatically change their well-being. But Brickman, Coates, and Janoff-Bulman’s study showed these predictions are actually wrong. Although winning the lottery was rated by its winners as a highly positive event, and accidents were rated by victims as a highly negative event, their happiness levels were remarkably close. In particular, lottery winners and the individuals in the control group “were not significantly different,” and while victims did report that they were less happy than before, they “did not appear nearly as unhappy as might have been expected.”

Against this background, recent legal scholarship has challenged the basis for victims’ right to hedonic damages (for an excellent discussion, see Bagenstos and Schlanger 2007). If victims adapt to their new physical condition even in the case of severe injury, the actual hedonic loss might be rather moderate, and the costs imposed on injurers are often excessive. Furthermore, the right itself might undermine victims’

adaptation process because it encourages victims to focus on their physical condition and the consequential losses rather than to rebuild their lives and regain their preinjury happiness level. Given judges' and jurors' failure to gauge the actual implications of disability, scholars have suggested abolishing hedonic damages altogether.

The behavioral argument against hedonic damages seems, however, to rely on a somewhat limited view of the psychological literature on adaptation. This literature emphasizes, as noted, that the effects of both negative and *positive events* (such as the award of money) gradually erode over the course of time. As Brickman, Coates, and Janoff-Bulman and other researchers showed, lottery winners' level of happiness spikes right after their winning but then progressively approaches their prewinning levels. This, in turn, suggests that tort compensation for nonmonetary harms might have a more moderate effect than conventional wisdom maintains. Accordingly, the findings on individuals' adaptation do not only suggest that victims' hedonic harm may be smaller than what is assumed by judges, but also that monetary compensation does not produce so much happiness as judges predict. As victims quickly "adapt" to the money they receive for their hedonic loss, the effect of such compensation on their well-being gradually erodes. Paradoxically, from a normative perspective, this erosion may call for either of two opposite conclusions. On the one hand, this erosion balances victims' adaptation to harm, which means that pain and suffering damages should be higher than what the current literature on hedonic adaptation suggests.<sup>12</sup> On the other hand, if damage awards are badly eroded, there may be a case for abandoning them altogether. Without further research, it is thus hard to determine, in the final tally, whether or not the current amounts for hedonic harm, and for pain and suffering in general, are adequate.

#### 2.4.4 *Punitive Damages*

As opposed to conventional tort payments, punitive damages are not intended to make victims whole. As their name implies, they are payments that come on top of the standard compensatory damages injurers pay in everyday tort lawsuits. When punitive damages are awarded, injurers' liability thus seems to exceed the social costs of their behavior.

Law and economics has offered several rationales to justify the imposition of above-compensatory damages. Most notable is the argument that punitive damages are necessary to attain optimal deterrence in a world in which enforcement against tortfeasors is less than 100% (Shavell 2004). Although it is usually in victims' interest to sue their injurers, in actuality they may nevertheless give up on meritorious claims. This may occur, for example, in cases in which victims cannot prove essential elements of their claim, when litigation costs exceed victims' expected compensation, or when victims prefer not risking their relationships with the injurer (e.g., workers and patients injured by their employers and doctors, respectively). Punitive damages serve to charge such injurers, once they are sued, for the loss in those cases in which they evaded

<sup>12</sup> We thank Alon Harel for this comment.

liability. Thus, according to this argument, punitive damages do not make injurers pay excessive damages but rather align their liability with the actual social costs of their behavior.

Experiments focusing on punitive damages, however, have provided little support for this suboptimal enforcement argument. Sunstein, Schkade, and Kahneman (2000) presented several groups of participants with hypothetical cases involving an injurer whose negligent activity harmed a victim. The hypothetical scenarios were identical in all aspects (the behavior of the injurer, victims' harm, etc.) except for the probability of litigation. In each scenario, participants were explicitly informed of the likelihood that the victim would sue. Participants were then requested to determine the appropriate compensation. The results showed that the probability of litigation had almost no effect on participants' answers. Although the instructions that were provided to the participants explicitly alluded to deterrence considerations, a low probability of litigation did not induce participants to award the victim higher compensation.

Looking to shed further light on individuals' perception of punitive damages, Sunstein and his colleagues conducted a follow-up experiment that directly focused on the interplay between optimal enforcement and punitive damages. In this study, participants were requested to evaluate a judge's decision not to award punitive damages in a case in which the injurer's behavior was particularly outrageous. As they were told, the judge rejected the plaintiff's plea for punitive damages because the probability of detection in such cases approaches 100%, and thus optimal deterrence can be achieved by imposing compensatory damages only. Contrary to the judge's reasoning, a vast majority of the participants found the decision unconvincing and supported the plaintiff's plea for punitive damages.

As this follow-up experiment suggests, in the decision whether to impose punitive damages, participants are particularly concerned with the reprehensibility of injurers' behavior. Further experiments have shown that individuals indeed use punitive damages to convey their disapproval when injurers' conduct is particularly unacceptable (Sunstein, Kahneman, and Schkade 1998). The farther removed an injurer's conduct is from the social norm, the greater the indifference of the injurer to victims' interests, the more likely individuals are to impose high punitive damages. Moreover, these experiments have shown that individuals often share common moral views in appraising injurers' conduct. When requested to rank hypothetical cases by injurers' blameworthiness, participants of different backgrounds and locations presented a remarkable consensus.

Nevertheless, studies exploring mock jurors' decisions to award punitive damages show that this consensus unravels once participants need to determine the dollar amounts to be paid by injurers. Despite sharing similar moral intuitions, participants show great variance in translating their ethical perceptions to specific monetary values. The lack of guidance on how to convert moral judgments to dollar payments results in erratic and unpredictable punitive damage awards (Sunstein et al. 2003).

The difficulty of translating moral judgments to dollar amounts makes jurors vulnerable also to cognitive manipulation by the litigants. In a study by Hastie,

Schkade, and Payne (1999a), an identical tort scenario was presented to several groups of mock jurors. Participants were then reviewing the closing argument of the plaintiff's lawyer. In one condition of the experiment, the lawyer demanded a relatively low amount in punitive damages, in the range of "15 and 50 million dollars". In the other condition, the lawyer demanded a higher amount, in the range of "50 to 150 million dollars." Although participants were instructed that plaintiff's arguments are not evidence, the results showed a significant "anchoring" effect. Participants in the second condition awarded the plaintiff in punitive damages nearly 2.5 times as much as those in the first condition. (On the anchoring effect in the legal context, see, e.g., Robbennolt and Studebaker 1999.)

These behavioral findings suggest that punitive damages are a challenge for the legal system. In their current format, punitive damages induce arbitrary outcomes and are potentially subject to manipulations by the parties. Moreover, studies have shown that merely providing jurors with further instructions, or encouraging them to deliberate, does not resolve the problem (and may only make things worse) (Schkade, Sunstein, and Kahneman 2000).

### 3 UNDERRESEARCHED AREAS IN TORTS AND POSSIBLE FURTHER CONTRIBUTIONS OF BEHAVIORAL ECONOMICS

---

In this section we offer two suggestions for further research in the crossroads between tort law and behavioral economics. We do not suggest that the issues here are the only possible avenues, or even the primary ones. A deeper exploration of the various topics discussed above is still very much missing, and can significantly enhance our understanding of the tort system. We do, however, find that both suggested topics are especially interesting for future research, because of the surprisingly limited (if any) attention given to them by behavioral law scholars. This lack of attention is startling because of the extensive attention given to these subjects—vicarious liability and perception of group variability—by law and economics scholars on the one hand, and by behavioral scholars on the other.

#### 3.1 The Behavioral Economics of Vicarious Liability

##### *3.1.1 Vicarious Liability and Its Basic Economic Analysis in a Nutshell*

Victims who suffer harm usually do not limit their lawsuit to their direct injurer. The reason is that, unless they happen to have extensive insurance, direct injurers often lack sufficient resources to pay victims' high damages. Suing the direct injurer alone

frequently results in the latter's bankruptcy, which in turn leaves victims uncompensated. To avoid this outcome, victims commonly seek to sue their injurer's employer (when such exists), relying on the doctrine of "respondeat superior" (Keeton et al. 1984). This doctrine renders employers—whose liquidity is higher and who are often insured—liable for the negligence of their employees. Accordingly, whether victims can actually collect compensation often hinges upon their ability to show that their direct injurer was an agent (employee) acting on behalf of a principal (employer).

Although tort law extends employers' liability to the negligent act of those who work for them, it does so only as long as employers possess sufficient control over their agents. Tort law distinguishes between "workers" and "independent contractors." In this regard, courts have considered such factors as who decides the details of the job, who supplies the place and instrumentalities of the work, and the method of payment. Employees who have little discretion in how to carry out their duties, who do not own their tools, and whose payment is time-based, are assumed to be under their employers' control. Recognizing that such workers are in effect their employers' long arm, "respondeat superior" enables victims to sue these workers' employers through vicarious liability. In contrast, employees who have wide discretion, own their tools, and are paid a lump sum are assumed to be under their own control. Such "independent contractors" are thus exclusively liable for the harm they cause, and their negligence cannot be imputed to those who hired them (see generally Kraakman 2009).

Despite its intuitive appeal, this distinction between workers and independent contractors has been criticized by law and economics scholars (Arlen and MacLeod 2005). While vicarious liability aims to induce employers to exert control over workers, so the argument goes, in fact it incentivizes employers engaging in harmful activities to choose independent contractors over hiring in-house employees. Furthermore, the independent contractors hired will likely have limited assets, such that when harm occurs victims will receive no compensation. Therefore, rather than reducing the risk of harm and increasing compensation, respondeat superior decreases social welfare and makes victims worse off.

To see the intuition of this argument, consider an employer who seeks an agent to perform a job that may result in harm. From the employer's perspective, a (nearly insolvent) in-house worker means it will be liable for any harm caused by its worker. Hiring instead an independent contractor is not, in itself, a superior option—because the independent contractor will likely charge more to compensate for the damages it pays victims. But if the independent contractor is thinly capitalized, it faces no risk of liability. Such an independent contractor will thus be able to perform the job at the same price as an in-house employee (while releasing the employer from potential liability). Accordingly, precisely in the context in which vicarious liability is most important (insolvent agents) it is unlikely to induce optimal behavior.

### 3.1.2 *Possible Implications for Behavioral Economics*

The conventional law-and-economics argument highlights the costs that employers can save by choosing independent contractors over in-house employees. Emerging

psychological literature suggests that ethical considerations, too, may induce employers to favor independent contractors over in-house employees. As studies establish, individuals tend to perceive indirect causation of negative outcomes as less unethical than direct causation, even if the outcome itself is worse in absolute terms (see generally, Royzman and Baron 2002). As these studies further show, indirect causation is often associated with the delegation of an action to self-controlled agents. Accordingly, by taking an independent agent, employers can distance themselves from victims' harm.

Early evidence for the role of agents is Milgram's seminal experiments on obedience (Milgram 1974). Milgram showed that subjects were more likely to cause pain to an innocent victim (following the direction of an authoritative figure) when the act of inflicting harm was removed from the subject and delegated to another participant (a confederate). This result led Milgram to suggest that "any force that is placed between the subject and the consequences of shocking the victim, any factor that will create distance between the subject and the victim, will lead to a reduction of strain on the participant and thus lessen disobedience" (p. 121).

More recently, Paharia and colleagues (2009) examined more closely how the use of intermediators affects people's moral judgments. Their study shows that participants evaluated the behavior of injurers as more unethical when it caused harm directly, as compared to the case in which it caused *an even greater harm* but through an intermediary. Thus, a pharmaceutical company's decision to raise the price of its cancer-related drug from \$3 to \$9 was perceived as far less ethical than its decision to sell the rights to the drug to another company, which raised the price to \$15. Similarly, a real-estate company's decision to clean up only 40% of the toxic waste on the land of its new housing project was perceived as more unethical than its decision to sell the land to a lesser-known company that did not engage in any clean-up effort. And likewise, a company that owned a polluting factory and decided to increase the pollution by 25% was rated more unethical than the same company when it sold this factory to a lesser-known company that increased the pollution by 75%.

In further experiments, Paharia and colleagues examined the effects of explicitly informing participants about the principal's levels of knowledge and control. These experiments showed that the existence of an agent attenuates the moral condemnation for companies' behavior, but the "indirectness effect" was eroded when the company had foreknowledge of the agents' actions, and particularly when it became evident that the agent was in fact the principal's long arm. Once participants became aware of the principal's knowledge and control over the agent, the agent's behavior was largely imputed to the principal itself.

The psychological experiments thus bear on the conventional economic argument on the social undesirability of the distinction between workers and independent contractors. First, these studies suggest that not only economic incentives but also ethical-perception incentives encourage employers to prefer independent contractors to workers. Accordingly, while the economic argument calls for making employers pay for victims' harm in the case of insolvent independent contractors (to eliminate the



incentives to hire them *ex ante*), the behavioral studies show that it might not be sufficient. Injurers might prefer to pay relatively high compensation for harm caused by an independent contractor than a lower amount for harm caused by their own worker. The use of an independent contract allows employers, concerned with how they are perceived, to psychologically disassociate themselves from victims' harm.

Second, these behavioral findings point out that the incentives to hire independent contractors might exist irrespective of their solvency. Relying on independent contractors might be an effective way to protect employers' reputation. Although solvent independent contractors are likely to charge for their costs of liability, employers may nevertheless profit from using their services rather than relying on in-house workers. Avoiding control over the acts of their agents may enable employers to reduce the public condemnation for the harm caused by these agents' activity.

Finally, these studies suggest that employers can benefit from using agents to diffuse liability. As noted earlier, punitive damages are often imposed once injurers' behavior exceeds a certain ethical threshold that reflects an outright indifference to victims' rights. This implies that employers can benefit from hiring an independent contractor as a means to reduce the risk of punitive damages. By dividing the responsibility between themselves and their independent contractors, employers may ensure that each of them will be perceived by juries as only partly responsible for the harm, at a level below the critical threshold. Thus an activity that may trigger punitive damages if done by the injurer itself (or its employee) may lead to a different result if it is handed over to an independent contractor.

## 3.2 The Behavioral Economics of Victims' Heterogeneity

### 3.2.1 *Underestimation of Group Variability*

People often perceive groups as more homogeneous than they really are. In exploring this phenomenon, social psychologists have devoted extensive attention to the out-group homogeneity (OH) bias, or OH effect, namely, people's tendency to perceive a group to which they do not belong as more homogeneous than it really is, and as less variable than their own group (e.g., Judd et al. 2005; Boldry, Gaertner, and Quinn 2007).<sup>13</sup> Furthermore, the perception of a group as homogeneous induces collective treatment that, in turn, reinforces the perception that this group is homogeneous (Alter and Darley 2009). Surprisingly, however, legal scholars in general, and tort scholars in particular, have not explored the implications of this bias for tort law, and above all for accidents that involve multiple victims.

<sup>13</sup> Recent research has also found what may be described as the opposite effect: an in-group homogeneity effect (Rubin and Badea 2012). However, for reasons laid *infra*, in tort situations the OH effect seems much more relevant.



Although it is one of the origins of the stereotypization process, the OH bias was also found to be robust in controlled laboratory experiments where no social stereotypes could be found. As studies have shown, the OH bias exists so long as members of the in-group on the one hand, and members of the out-group on the other, share some minimal common traits—even artificial, laboratory-made characteristics (Rubin, Hewstone, and Voci 2001; Voci 2000). The strength and robustness of this bias are affected by several factors: the out-group size (the larger the group the larger the effect), the social power gap between the in-group and the out-group (the greater the gap between groups the larger the effect), and familiarity with members of the out-group (the more familiarity with group members the lesser the effect) (Rubin and Badea 2007, 2012).<sup>14</sup>

### 3.2.2 *Heterogeneity of Tort Victims*

In modern reality, accidents involving multiple (heterogeneous) victims are common: environmental hazards, mass production of dangerous products, the increasing power of machines and vehicles, the rise in population density, alongside the social, demographic, technological, and economic progressions of the modern era, all give rise to a growing number of multiple-victim accidents. It is a fact of life that parties to accidents are not identical to one another. This heterogeneity refers, first and foremost, to victims—even victims of the same accident. The level of risk to which each victim is subject, as well as the costs of precautionary measures available to each victim and their efficacy—all depend on factors that naturally vary between victims, such as their age, health condition, wealth, occupation, location, and the value of their property. When victims themselves engage in the risky activity (such as using risky products), the utility they derive from this activity may also vary considerably. Hence, the economic factors that determine the assignment of liability in multiple-victim torts—the harm and its *ex ante* probability, the costs of precautions, and the utility derived from activity—typically differ from one victim to another.

The features that affect the size and robustness of the OH bias—the out-group size, the social power gap between the in-group and the out-group, and unfamiliarity with members of the out-group—make tort situations especially vulnerable to it. Furthermore, experimental studies show that the mere effect of being harmed collectively, even by random events, induces a more homogeneous perception of the group of victims. For example, in one of their studies, Alter and Darley (2009) gave participants a fabricated newspaper article describing harm caused by a tornado. Participants, who were told that 40 apartments in a complex were harmed by a tornado, perceived the apartment-owners as more homogeneous in their nationalities, languages, and ethnicities than participants who were told that only one of the apartment-owners was affected by the tornado. As the authors note, because “a tornado does not selectively affect homogeneous groups of people, participants could

<sup>14</sup> Nevertheless, one study from the early 1980s (Jones, Wood, and Quattrone 1981) may suggest that unfamiliarity with out-group members is not paramount for perceived homogeneity.

not logically infer that the collective treatment followed from some preexisting difference in homogeneity that distinguished the two groups of people.” Hence, both injurers and the courts might perceive groups of victims as more homogeneous than they actually are.

The social psychology scholarship has not deeply explored the possibility of debiasing the OH effect. Although the literature has indeed explored the factors that affect the size and robustness of this bias, these factors relate to exogenic facts that are characteristics of the out-group. They cannot be “manipulated” in the real world in a way that would fit a debiasing effort, since exposure to the common issues among the group of victims—whatever enhancing effect it may have on the OH bias—is nonetheless essential for procedural matters such as the certification of groups in class actions. It is also obvious that in the legal context, requiring the court to have personal familiarity with members of the group of victims is both unrealistic and contrary to the basic principles of justice.

### 3.2.3 *Legal Applications: Standard of Care with Multiple Victims*

The long-standing conventional wisdom in torts maintains that, when an injurer inflicts a risk of harm on more than one victim, this injurer’s liability toward the victims should be based on an aggregative cost-benefit test, which weighs the accident’s overall social costs against the overall social benefits of the injurer’s conduct. This aggregative test is also relevant for determining the victims’ contributory negligence in multiple-victim torts (Halbersberg 2013; Restatement (Third) of Torts: American Law Institute 2010, §3 cmt. e). However, when victims are heterogeneous, this aggregative test may be ill suited to the determination of liability (Halbersberg 2013). Take, for example, a case where the injurer’s activity inflicted harm on only two victims, whose precaution costs and expected harm differs as shown in table 16.1.

Applying an overall risk-benefit test will find the injurer liable, because she could have prevented an overall risk of 70 had she invested only 50 in precautions, while the victims themselves could have prevented the harm only at a cost of 80. A closer look, however, reveals that a more efficient outcome would be reached if the injurer were exempt from liability, thereby inducing Victim 1 to take precautions and Victim 2 to bear the harm. If this were the case, the overall social costs would amount to only 40—less than the injurer’s precaution costs of 50.<sup>15</sup>

This efficient outcome could be reached only if courts were sensitive to the victims’ heterogeneity, and used a standard of liability that aggregated, for each victim, the lesser of either the expected harm or costs of precautions and compared this amount to the injurer’s costs (see further, Halbersberg 2013). However, such a nuanced approach

<sup>15</sup> This is case where the conventional test leads to overdeterrence of the injurer. Halbrsberg (2013) lists other categories of cases where heterogeneity of victims makes the conventional test for negligence, for contributory negligence, or for design defects in product liability cases inefficient, including underdeterrence and dilution of liability.

**Table 16.1 Real Costs of Accident**

	Victim 1	Victim 2	Aggregate Victim	Injurer
Expected harm	60	10	70	
Precaution costs	30	50	80	50

**Table 16.2 Perceived Costs of Accident with OH bias**

	Victim 1	Victim 2	Aggregate Victim	Injurer
Expected harm	35	35	70	
Precaution costs	30	30	60	50

that is attuned to the differences between victims is unlikely under the OH bias. The effect of the OH bias is to “average out” victims’ differences and to make them be perceived as more homogeneous than they really are. The OH bias will cause the injurer, as well as the court, to assume that victims’ expected harm and costs of precautions are as shown in table 16.2.

One can readily see the implications of this change in the perception of victims’ homogeneity. Under a negligence standard, if either the injurer or the fact-finder were affected by the OH bias, then the injurer would be driven to take the socially inefficient precautions, leading to a social loss, in this case, of 10 ( $50 - 40$ ).

Unfortunately, experiments that tested tort liability for environmental disasters (Angelova, Attanasi, and Hiriart 2012) did not explore the effects of the OH bias and were designed as if one victim suffered the large harm in its entirety. Alter and Darley (2009) did demonstrate that victims are perceived as more homogeneous, but did so in a context unrelated to tort liability.

Despite the lack of experimental work that examines the implications of the OH bias for mass tort cases, there seems to be a consensus among legal practitioners that juries tend to perceive victims as less variable than they actually are—so that, in the juries’ minds, tort victims with high levels of harm “balance” other victims who suffer less harm, and an averaging effect is apparent (e.g., McNeil and Fancsali 1996, p. 491). There are court decisions that suggest that some homogeneity bias may indeed have been present. One such example is the Israeli case of *Haifa Municipality*.<sup>16</sup> This case originated from four separate events of flooding that caused harm to a heterogeneous group of 21 property owners, including factories, small businesses, and private individuals. Despite their heterogeneity, after describing the idiosyncratic facts surrounding only one case out of the twenty-one,<sup>17</sup> the court succinctly determined that *all* victims had the same 20% share of fault under the Israeli comparative negligence

<sup>16</sup> AC 2906/01 Haifa Municipality v. Menorah Insurance Company.

<sup>17</sup> In one factory, the damaged goods were placed on the floor after the first flood, making them vulnerable to being harmed in later floods.

rule. On appeal, the Israeli Supreme Court agreed. It can be argued, therefore, that this case is an example of the perception of multiple victims as homogeneous, and the equal treatment that was consequentially accorded to them.

The effect of OH bias is not limited to the proper determination of parties' negligence. First, as already mentioned, group variability influences the litigation of mass torts. The more homogeneous the group of victims is perceived to be, the more likely courts are to certify pleas for class action. Second, the use of bellwether trials<sup>18</sup> may be more acceptable if the victims group is perceived as more homogeneous than it really is. Third, victims' heterogeneity can greatly affect the efficiency of, and justification for, applying the doctrine of market share liability (MSL) (see generally Gifford and Pasicolan 2006). This doctrine allows a group of victims—who can point to a group of injurers as a whole as causing harm, but are incapable of proving which victim was harmed by which injurer—to nevertheless recover the harm in proportion to each injurer's market share. However, both the efficiency and moral justifications for this doctrine require that the victims' group be homogeneous. If victims are heterogeneous but are perceived as homogeneous, courts might use MSL even though its justifications do not apply in the particular case. Furthermore, the basic economic justification for punitive damages, as a realignment of injurers' liability with the actual social costs of their behavior, as well as the conventional formula for their calculation, also requires that the victims be fairly homogeneous. Recall that the conventional formula maintains that, to achieve optimal deterrence, the overall damages (punitive + compensatory) should equal the compensatory damages multiplied by the inverse of the probability of being found liable (Polinsky and Shavell 1998). But when the group of victims is heterogeneous, the determination of punitive damages requires a more idiosyncratic calculation. Consequently, if the victims are perceived as more homogeneous than they really are, this may impinge on the overall efficiency of awarding punitive damages.

The exploration of the OH bias (as well as other social-psychology biases that affect the perception of victims' heterogeneity) thus seems a promising avenue for future studies. Further research may examine how the OH bias actually affects juries, judges, and injurers as well as potential tools the legal system could employ to induce such parties to identify differences between victims of harmful activities. Further behavioral research may also wish to explore possible ways for debiasing the OH bias in general, and in the legal context in particular—such as using small samples from the victims' group.

<sup>18</sup> A bellwether trial is a process by which “a random sample of cases large enough to yield reliable results is tried to a jury (in order to) use the resulting verdicts as a basis for resolving the remaining cases” (Lahav 2008, p. 577).

## 4 CONCLUSION

---

As this chapter suggests, behavioral studies have already provided significant contributions to the understanding of the actual implications of our tort system. These studies, as we have shown, shed light on many fundamental questions regarding the proper regulation of harmful activities and the optimal design of victims' right to compensation. The fast-growing number of experimental studies that focus on legal issues in general, and tort-related contexts in particular, promises that our chapter will likely be in a need of expansion (and revision) in the coming future.

### REFERENCES

---

- Adler, M., and E. A. Posner. 1999. Rethinking Cost-Benefit Analysis. *Yale Law Journal* 109:165–247.
- Alter, A. L., and J. M. Darley. 2009. When the Association between Appearance and Outcome Contaminates Social Judgment: A Bidirectional Model Linking Group Homogeneity and Collective Treatment. *Journal of Personality and Social Psychology* 97:776–95.
- American Law Institute. 2010. *Restatement of the Law Third, Torts—Liability for Physical and Emotional Harm*. St. Paul, MN: American Law Institute Publishers.
- Angelova, V., G. Attanasi, and Y. Hiriart. 2012. Relative Performance of Liability Rules: Experimental Evidence. TSE Working Papers 12-304, Toulouse School of Economics, revised September.
- Arlen, J., and W. B. Macleod. 2005. Beyond Master-Servant: A Critique of Vicarious Liability. In M. Stuart Madden (ed.), *Exploring Tort Law*. New York: Cambridge University Press, pp. 111–42.
- Avraham, R. 2005. Should Pain-and-Suffering Damages Be Abolished from Tort Law?: More Experimental Evidence. *University of Toronto Law Journal* 55 (Fall): 941–80.
- Bagenstos, S. R., and M. Schlanger. 2007. Hedonic Damages, Hedonic Adaptation, and Disability. *Vanderbilt Law Review* 60:745–97.
- Baron, J., and J. C. Hershey. 1988. Outcome Bias in Decision Evaluation. *Journal of Personality and Social Psychology* 54:569–79.
- Bhole, B., and J. Wagner. 2010. Punitive Damages and the Recklessness Requirement with Uninformed Injurers. *International Review of Law and Economics* 30:253–64.
- Bigus, J. 2006. Tort Liability and Probability Weighting Function According to Prospect Theory. Paper presented at the Sixteenth Annual Meeting of the American Law and Economics Association, Berkeley, CA. Available at <http://law.bepress.com/cgi/viewcontent.cgi?article=1694&context=alea>.
- Boldry, J. G., L. Gaertner, and J. Quinn. 2007. Measuring the Measures: A Meta-analytic Investigation of the Measures of Outgroup Homogeneity. *Group Processes and Intergroup Relations* 10:157–78.
- Brickman, P., D. Coates, and R. Janoff-Bulman. 1978. Lottery Winners and Accident Victims: Is Happiness Relative? *Journal of Personality and Social Psychology* 36:917–27.
- Brown, J. P. 1973. Toward an Economic Theory of Liability. *Journal of Legal Studies* 2:323–50.

- Calabresi, G. 1961. Some Thoughts on Risk Distribution and the Law of Torts. *Yale Law Journal* 70:35–46.
- Calabresi, G. 1970. *The Cost of Accidents: A Legal and Economic Analysis*. New Haven, CT: Yale University Press.
- Calabresi, G., and J. T. Hirschoff. 1972. Toward a Test for Strict Liability in Tort. *Yale Law Journal* 81:1055–85.
- Camerer, C.F. 1990. Comment on Noll and Krier, “Some Implications of Cognitive Psychology for Risk Regulation”. *Journal of Legal Studies* 19:791–92.
- Cooter, R., and T. S. Ulen. 2012. *Law and Economics*. 6th ed. Boston: Addison-Wesley.
- Coase, R. 1960. The Problem of Social Cost. *Journal of Law and Economics* 3:1–44.
- Crane, D. A. 2011. Rethinking Merger Efficiencies. *Michigan Law Review* 110:347–391.
- Diener, E., and C. Diener. 1996. Most People Are Happy. *Psychologic Science* 7:181–85.
- Faure, M. G. 2010. Behavioral Accident Law and Economics. *Journal of Applied Economy* 4:11–68.
- Feldman, Y., and D. Teichman. 2008. Are All Legal Dollars Created Equal? *Northwestern University Law Review* 102:223–62.
- Fischhoff, B. 1975. Hindsight [Not =] Foresight: The Effect of Outcome Knowledge on Judgment under Uncertainty. *Journal of Experimental Psychology: Human Perception and Performance* 1:288–99.
- Frederick, S., and G. Loewenstein. 1999. Hedonic Adaptation. In Kahneman, Daniel, E. Diener, and N. Schwarz (eds.), *Well-being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation, pp. 302–29.
- Gifford, G., and P. Pasicolan. 2006. Market Share Liability beyond DES Cases: The Solution to the Causation Dilemma in Lead Paint Litigation?. *South Carolina Law Review* 58:115–59.
- Ghosh, S., and R. P. Kundu. 2010. Efficiency of Liability Rules: An Experimental Analysis. Available at SSRN: <http://ssrn.com/abstract=1684009>.
- Gilo, D., and E. Guttel. 2009. Negligence and Insufficient Activity: The Missing Paradigm in Torts. *Michigan Law Review* 108:277–322.
- Guttel, E., and A. Harel. 2005. Matching Probabilities: The Behavioral Law and Economics of Repeated Behavior. *University of Chicago Law Review* 72:1197–236.
- Halbersberg, Y. 2013. Toward a New Paradigm for Multiple-Victim Torts: The Problem of Victims’ Heterogeneity. <http://ssrn.com/abstract=1264464>.
- Hamilton, J. T., and W. K. Viscusi. 1999. Are Risk Regulators Rational? Evidence from Hazardous Waste Cleanup Decisions. *American Economic Review* 89:1010–27.
- Harris, R., and E. Albin. 2006. Bankruptcy Policy in Light of Manipulation in Credit Advertising. *Theoretical Inquiries in Law* 7:431–65.
- Hastie, R., D. A. Schkade, and J. W. Payne. 1999a. Juror Judgments in Civil Cases: Effects of Plaintiff’s Requests and Plaintiff’s Identity on Punitive Damage Awards. *Law and Human Behavior* 23:445–70.
- Hastie, R., D. A. Schkade, and J. W. Payne. 1999b. Juror Judgments in Civil Cases: Hindsight Effects on Judgments of Liability for Punitive Damages. *Law and Human Behavior* 23:597–614.
- Hastie, R., and W. K. Viscusi. 1998. What Juries Can’t Do Well: The Jury’s Performance as a Risk Manager. *Arizona Law Review* 40:901–21.
- Hastie, R., and W. K. Viscusi. 2002. Juries, Hindsight, and Punitive Damages Awards: Reply to Richard Lempert. *DePaul Law Review* 51:987–98.
- Issacharoff, S. 2002. Preclusion, Due Process, and the Right to Opt Out of Class Actions. *Notre Dame Law Review* 77:1057–81.



- Jolls, C., C. R. Sunstein, and R. Thaler. 1998. A Behavioral Approach to Law and Economics. *Stanford Law Review* 50:1471–550.
- Jones, E. E., G. C. Wood, and G. A. Quattrone. 1981. Perceived Variability of Personal Characteristics in In-Groups and Out-Groups: The Role of Knowledge and Evaluation. *Personality and Social Psychology Bulletin* 7:523–28.
- Judd, C. M., B. Park, V. Yzerbyt, E. Gordijn, and D. Muller. 2005. Attributions of Intergroup Bias and Outgroup Homogeneity to Ingroup and Outgroup Others. *European Journal of Social Psychology* 35:677–704.
- Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47: 63–92.
- Kamin, K. A., and J. J. Rachlinski. 1995. Ex Post Ex Ante: Determining Liability in Hindsight. *Law and Human Behavior* 19:89–103.
- Keeton, W. P., D. B. Dobbs, R. E. Keeton, and D. G. Owen. 1984. *Prosser and Keeton on the Law of Torts*. 5th ed. St. Paul, MN: West.
- Kornhauser, L., and A. Schotter. 1990. An Experimental Study of Single-Actor Accidents. *Journal of Legal Studies* 19:203–33.
- Kornhauser, L., and A. Schotter. 1991. An Experimental Study of Two-Actor Accidents. Working Papers 91-60, C.V. Starr Center for Applied Economics, New York University.
- Korobkin, R. B., and T. S. Ulen. 2000. Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics. *California Law Review* 88:1051–144.
- Kraakman, R. H. 2009. Vicarious and Corporate Liability. In M. Faure (ed.), *Tort Law and Economics*. 2nd ed. Northampton, MA: Edward Elgar, pp. 669–81.
- LaBine, S. J., and G. LaBine. 1996. Determinations of Negligence and the Hindsight Bias. *Law and Human Behavior* 20:501–16.
- Lahav, A. 2008. Bellwether Trials. *The George Washington Law Review* 76:576–638.
- Landes, W. M., and R. A. Posner. 1987. *The Economic Structure of Tort Law*. Cambridge, MA: Harvard University Press.
- Lempert, R. 1999. Juries, Hindsight, and Punitive Damage Awards: Failures of a Social Science Case for Change. *DePaul Law Review* 48:867–94.
- McCaffery, E. J., D. J. Kahneman, and M. L. Spitzer. 1995. Framing the Jury: Cognitive Perspectives on Pain and Suffering Awards. *Virginia Law Review* 81:1341–420.
- McNeil, B. F., and B. L. Fancsali. 1996. Mass Torts and Class Actions: Facing Increased Scrutiny. *Federal Rules Decisions* 167:483–522.
- Milgram, S. 1974. *Obedience to Authority: An Experimental View*. New York: HarperCollins.
- Molot, J. T. 1997. How U.S. Procedure Skews Tort Law Incentives. *Indiana Law Journal* 73: 9–117.
- Note. 1993. Settling for Less: Applying Law and Economics to Poor People. *Harvard Law Review* 107:442–59.
- Noll, R., and J. Krier. 1990. Some Implications of Cognitive Psychology for Risk Regulation. *Journal of Legal Studies* 19:747–79.
- Paharia, N., K. S. Kassam, J. D. Greene, and M. H. Bazerman. 2009. Dirty Work, Clean Hands: The Moral Psychology of Indirect Agency. *Organizational Behavior and Human Decision Processes* 109:134–41.
- Peters, P. G., Jr. 1999. Hindsight Bias and Tort Liability: Avoiding Premature Conclusions. *Arizona State Law Journal* 31:1277–314.
- Polinsky, A. M., and S. Shavell. 1998. Punitive Damages: An Economic Analysis. *Harvard Law Review* 111:869–962.



- Porat, A., and A. Stein. 2001. *Tort Liability under Uncertainty*. Oxford: Oxford University Press.
- Poser, S., B. H. Bornstein, and E. K. McGorty. 2003. Measuring Damages for Lost Enjoyment of Life: The View from the Bench and the Jury Box. *Law and Human Behavior* 27:53–68.
- Posner, E. A. 2004. Probability Errors: Some Positive and Normative Implications for Tort and Contract Law. *Supreme Court Economic Review* 11:125–41.
- Posner, R. A. 1972a. A Theory of Negligence. *Journal of Legal Studies* 1:29–96.
- Posner, R. A. 1972b. *Economic Analysis of Law*. Boston: Little, Brown.
- Posner, R. A. 1973. Strict Liability: A Comment. *Journal of Legal Studies* 2:205–21.
- Posner, R. A. 2007. *Economic Analysis of Law*. 7th ed. New York: Aspen.
- Rachlinski, J. J. 1998. A Positive Psychological Theory of Judging in Hindsight. *University of Chicago Law Review* 65:571–625.
- Rachlinski, J. J. 1999. Regulating in Foresight versus Judging Liability in Hindsight: The Case of Tobacco. *Georgia Law Review* 33:813–45.
- Robbennolt, J. K., and C. A. Studebaker. 1999. Anchoring in the Courtroom: The Effects of Caps on Punitive Damages. *Law and Human Behavior* 23:353–73.
- Royzman, E. B., and J. Baron. 2002. The Preference for Indirect Harm. *Social Justice Research* 15:165–84.
- Rubin, M., and C. Badaea. 2007. Why Do People Perceive In-Group Homogeneity on In-Group Traits and Out-Group Homogeneity on Out-Group Traits? *Personality and Social Psychology Bulletin* 33:31–42.
- Rubin, M., and C. Badaea. 2012. “They’re All the Same!”. . . but for Several Different Reasons: A Review of the Multicausal Nature of Perceived Group Variability. *Current Directions in Psychological Science* 21:367–72.
- Rubin, M., M. Hewstone, and A. Voci. 2001. Stretching the Boundaries: Strategic Perceptions of Intragroup Variability. *European Journal of Social Psychology* 31:413–29.
- Schaefer, H.-B., and F. Mueller-Langer. 2008. Strict Liability versus Negligence. Available at SSRN: <http://ssrn.com/abstract=2062787>.
- Schäfer, H.-B., and A. Schönenberger. 2000. Strict Liability versus Negligence. In B. Bouckaert and G. De Geest (eds.), *Encyclopedia of Law and Economics*, vol. 2. Cheltenham: Edward Elgar, pp. 597–624.
- Schkade, D., C. R. Sunstein, and D. Kahneman. 2000. Deliberating about Dollars: The Severity Shift. *Columbia Law Review* 100:1139–76.
- Shavell, S. 1980. Strict Liability versus Negligence. *Journal of Legal Studies* 9:1–25.
- Shavell, S. 1984. Liability for Harm versus Regulation of Safety. *Journal of Legal Studies* 13:357–74.
- Shavell, S. 1987. *Economic Analysis of Accident Law*. Cambridge, MA: Harvard University Press.
- Shavell, S. 2004. *Foundations of Economic Analysis of Law*. Cambridge, MA: Harvard University Press.
- Shavell, S. 2013. A Fundamental Enforcement Cost Advantage of the Negligence Rule over Regulation. Available at <http://www.law.uchicago.edu/files/files/Shavell%20paper.pdf>.
- Stensman, R. 1985. Severely Mobility-Disabled People Assess the Quality of Their Lives. *Scandinavian Journal of Rehabilitation Medicine* 17:87–99.
- Sugden, R. 2002. Alternatives to Expected Utility: Foundations. In P. J. Hammond, S. Barberá, and C. Seidl (eds.), *Handbook of Utility Theory*. vol. 2. Boston, MA: Kluwer.
- Sunstein, C. R. 2000. Cognition and Cost-Benefit Analysis. *Journal of Legal Studies* 29:1059–103.

- Sunstein, C. R., R. A. Hastie, J. W. Payne, D. A. Schkade, and W. K. Viscusi. 2003. *Punitive Damages: How Juries Decide*. Chicago: University of Chicago Press.
- Sunstein, C. R., D. Kahneman, and D. A. Schkade. 1998. Shared Outrage and Erratic Awards: The Psychology of Punitive Damages. *Journal of Risk and Uncertainty* 16:49–86.
- Sunstein, C. R., D. A. Schkade, and D. Kahneman. 2000. Do People Want Optimal Deterrence?. *Journal of Legal Studies* 29:237–53.
- Teitelbaum, J. C. 2007. A Unilateral Accident Model under Ambiguity. *Journal of Legal Studies* 36:431–77.
- Viscusi, W. K. 1999. Using Warnings to Extend the Boundaries of Consumer Sovereignty. *Harvard Journal of Law and Public Policy* 23:211–31.
- Voci, A. 2000. Perceived Group Variability and the Salience of Personal and Social Identity. *European Review of Social Psychology* 11:177–221.
- Wittman, D. 1977. Prior Regulation versus Post Liability: The Choice between Input and Output. *Journal of Legal Studies* 6:193–211.
- Wittman, D., D. Friedman, S. Crevier, and A. Braskin. 1997. Learning Liability Rules. *Journal of Legal Studies* 26:145–64.
- Zamir, E., and I. Ritov. 2011. Notions of Fairness and Contingent Fees. *Law and Contemporary Problems* 74:1–32.

## CHAPTER 17

---

# BEHAVIORAL ECONOMICS AND CONTRACT LAW

---

MELVIN A. EISENBERG

## 1 INTRODUCTION

---

### 1.1 Formalism and Classical Contract Law

From the middle of the nineteenth century through the early part of the twentieth, contract theory was dominated by a school of thought now known as classical contract law. Theories of contract law fall into several broad categories, based on the method of legal reasoning utilized by theories in the category and the roles those theories assign to doctrinal and social propositions. (By doctrinal propositions, I mean propositions that purport to state legal rules and are found in or can be derived from sources that are generally regarded by the legal profession as authoritative. By social propositions, I mean moral norms, which characterize conduct as right or wrong; policies, which characterize states of affairs as good or bad, depending on whether they are conducive or adverse to the general welfare; and experiential propositions, which characterize the way in which the world works, and mediate between moral and policy propositions, on the one hand, and doctrinal propositions, on the other.) One category of contract theory is formalism, and classical contract law falls into this category. Formalism generally, and classical contract law in particular, are characterized by an ideological aspect, a methodological aspect, and a psychological aspect.

Ideologically, formalism and classical contract law conceived of legal doctrine as autonomous from policy and morality. This autonomy is well illustrated by a passage from *Summary of the Law of Contracts* by Christopher Langdell, a key figure in classical contract law, addressing the question whether an acceptance by mail is effective on acceptance or dispatch:

The acceptance. . . must be communicated to the original offerer, and until such communication the contract is not made. It has been claimed that the purposes of substantial justice, and the interests of contracting parties as understood by themselves, will be best served by holding that the contract is complete the moment the letter of acceptance is mailed; and cases have been put to show that the contrary view would produce not only unjust but absurd results. The true answer to this argument is that it is irrelevant. (Langdell 1880, 15, 20–21)

Methodologically, formalism and classical contract law proceeded in a Euclidean manner, by first identifying a set of postulates or general principles that were justified axiomatically on the ground that they were self-evident, and then deriving more specific rules by logical deduction from the axioms. As Holmes observed at the height of classical contract law, “I sometimes tell students that the law schools pursue an inspirational combined with a logical method, that is, the postulates are taken for granted upon authority without inquiry into their worth, and then logic is used as the only tool to develop the results” (Holmes 1920, 238). In a sense, formalism, and therefore classical contract law, had to adopt this methodology. If doctrine was to be autonomous from social propositions, the only path on which legal reasoning could proceed was the adoption of principles claimed to be self-evident and rules formed by deduction from those principles. The hermetic nature of this methodology was well put by James Bradley Thayer, commenting on an opinion of Chief Justice Holt:

The Chief Justice here retires into that lawyer’s Paradise where all words have a fixed, precisely ascertained meaning; where men may express their purposes, not only with accuracy, but with fullness; and where, if the writer has been careful, a lawyer, having a document referred to him, may sit in his chair, inspect the text, and answer all questions without raising his eyes. Men have dreamed of attaining for their solemn muniments of title such an absolute security; and some degree of security they have encompassed by giving strict definitions and technical meanings to words and phrases, and by rigid rules of construction. But the fatal necessity of looking outside the text in order to identify persons and things, tends steadily to destroy such illusions and to reveal the essential imperfection of language, whether spoken or written. (Thayer 1898, 428–29)

Psychologically, formalism and classical contract law were implicitly based on a rational-choice, or rational-actor, view of decision-making. Under this view, actors make choices that will maximize their utility based on a rational analysis of the expected outcomes of competing choices. Rationality, for this purpose, requires that an actor possesses and accurately processes all readily available information concerning the outcomes of alternative choices, ranks the possible outcomes in order of their expected utility, understands probability, and properly discounts future states of the world when comparing them with present states of the world.

Formalist legal theories, such as classical contract law, are untenable. Determining general principles axiomatically is untenable because doctrinal propositions are never self-evident, and can be justified only by social propositions. A distinction must be drawn here between the justification for *following* a doctrine and the justification of a

doctrine. For example, although a common-law court may justify a result on the basis of a rule stated in precedents, such a rule can always be traced back to a point where precedential justification is unavailable, and only a social justification will do. Once a doctrine has been adopted, it may be justifiably followed because of the social interest in stability and predictability, the social reasons for adhering to rules that have been adopted through certain institutional processes, or both. However, those elements only justify following a doctrine, not the doctrine itself. A doctrine itself is justified only if it rests on social propositions that are appropriate for generating legal doctrine. For example, at a time when most states had adopted the doctrine of contributory negligence, a trial court might have been justified in following that doctrine, but that would have little or no bearing on whether the doctrine itself was justified. The doctrine itself could be justified only on basis of moral, policy, and experimental propositions.

Determining specific legal rules by logical deduction from axiomatic general principles is no more sustainable than determining general principles on axiomatic grounds. Since a legal principle cannot be justified axiomatically, neither can a rule derived by logical deduction from a principle that is justified only axiomatically. Moreover, any given principle can be expressed through various rules, depending largely on what distinctions should be drawn in applying the principle; and what distinctions should be drawn depends on social propositions, not deductive logic. For example, suppose that precedent in a jurisdiction has adopted the principle that bargain promises are enforceable according to their terms in the absence of fraud, duress, or the like. The jurisdiction has not yet established whether age-based incapacity is a defense to a contract. Now an actor who made a bargain with a fourteen-year-old seeks to enforce the bargain. If common-law rules were fixed, then as a matter of logical deduction the minor would be liable. The major premise would be that in the absence of fraud, duress, or the like, bargains are enforceable according to their terms. The minor premise would be that the fourteen-year-old made a bargain. The conclusion would be that the fourteen-year-old is liable. But this conclusion should not and will not be drawn, because the social propositions that support the bargain principle do not support the application of that rule to persons whom society considers underage. One social reason for the bargain principle is that actors are normally the best judges of their own interests. Based on experience, this reason does not apply to persons society considers underage. Therefore, the principle should be reformulated by creating an exception for underaged persons.

In contrast, suppose a court was asked to hold that a bargain made by a clergyman is unenforceable against the clergyman even though the bargain was not religious in nature (that is, did not concern an issue of dogma, the allocation of authority within a church, or the like). This exception should not be engrafted onto the rule, because no applicable social proposition supports a special status for clergymen who make contracts. It is easy to imagine social propositions that would support a special status for clergymen for other purposes or in another society or time. For example, in the Middle Ages there was a clergyman exception under the criminal law—clergymen could be prosecuted for felony only in ecclesiastical courts and therefore were not subject to capital punishment (Pollock and Maitland 1898, 411–57). Even today, religious bargains

made by clergymen might well be unenforceable. But applicable social propositions in contemporary society would not support a clergyman exception to the bargain principle. In short, any common-law principle may be reformulated at any time by the processes of distinguishing and drawing exceptions. Whether and how a given rule should be reformulated in this way depends on social propositions.

The psychological basis of classical contract law is also open to serious question. Of course, actors ordinarily want to maximize their utility. The problems are whether actors are fully rational in processing information about alternative choices, discounting future states of the world and comparing them with present states of the world, understanding probability, and utilizing heuristics, that is, decision-making formulas or shortcuts.

## 1.2 Modern Contract Law

Beginning in the late 1920s and early 1930s, under the influence of Arthur Corbin, Karl Llewellyn, and other leading scholars, classical contract law was largely although not entirely supplanted by modern contract law. Where classical contract law is formal, modern contract law is substantive, that is, based on social propositions rather than axioms, logical deduction, and autonomous reasoning. So, for example, while classical contract law took the position, on purely axiomatic grounds, that only bargain contracts were enforceable, modern contract law takes the position, on grounds of policy and morality, that reliance makes a promise enforceable, at least to the extent of reliance. Even *Restatement First*, authored principally by Williston, often thought of as a central formalist text, published in 1932, adopted important new doctrines that were justifiable on substantive rather than formal grounds and *Restatement Second*, published in 1981, carried this development much further.

## 2 CHICAGO LAW AND ECONOMICS

---

In the 1970s and 1980s, as modern contract law was still emerging, two important social-science-based scholarly developments began to emerge as well. One of these was law and economics. Contract law concerns exchange, and therefore was long implicitly or explicitly intertwined with economic analysis. Like classical contract law, classical economics centers on formal reasoning (although mathematical rather than Euclidean) and a rational-actor model of psychology. Accordingly, although classical contract law did not explicitly incorporate classical economics, it mapped very well onto classical economics, and during the period when the school of classical contract law held sway, that school and classical economics proceeded along more or less parallel tracks.

In its inception, law-and-economics differed from classical economics in two respects. First, while classical economics mapped on to classical contract law, it was not about law. In contrast, law-and-economics was explicitly about law. Second, early law-and-economics was principally birthed by Richard Posner and was based on Chicago economics—markets are good and regulation isn't. The methodologies of Chicago law-and-economics and modern contract law are congruent insofar as both were ultimately based on social policy, but diverged insofar as Chicago law-and-economics was based on formal reasoning and rational-actor psychology. In effect, Chicago law-and-economics was a type of neoclassical contract law.

### **3 BEHAVIORAL PSYCHOLOGY AND BEHAVIORAL LAW AND ECONOMICS**

---

Around the time that law-and-economics began to emerge there also emerged a paradigm shift in the psychology of decision-making. This shift, which can be broadly referred to as behavioral psychology, has been manifested in three waves of scholarship. The first wave shows that actors often make decisions without having full information, without adequately processing the information they do have, without bringing to their conscious minds all the critical assumptions that underlie their decisions, or all three. The second wave shows that in certain areas actors systematically make decisions that are not rational. The third wave bears principally on how contracting actors behave—for example, how contracting actors think about sanctions for breach, are incentivized or disincentivized by given types of contractual provisions, respond to the way in which issues are framed, and so forth.

### **4 THE FIRST WAVE: BOUNDED RATIONALITY, RATIONAL IGNORANCE, AND TACIT ASSUMPTIONS**

---

#### **4.1 Bounded Rationality**

A basic assumption of classical contract law, classical economics, and neoclassical contract law is that contracts are made by informed parties. As Michael Trebilcock has pointed out:

Even the most committed proponents of free markets and freedom of contracts recognize that certain information preconditions must be met for a given exchange



to possess Pareto superior qualities. For example, to recall the statement of Milton Friedman: “The possibility of coordination through voluntary cooperation rests on the elementary—yet frequently denied—proposition that both parties to an economic transaction benefit from it, *provided the transaction is bilaterally voluntary and informed.*” (Trebilcock 1993, 117–18, quoting Milton Friedman 1962).

However, as shown by Herbert Simon and others in work beginning in the 1950s and 1960s, actors are almost never fully informed when they make decisions. If the costs of searching for and processing information were zero, and human information-processing capabilities were perfect, then an actor who wanted to make a choice would make a comprehensive search for all relevant information, would perfectly process the information he acquired, and would then make the best possible choice—the choice that, as of the time made, was better than all the alternative choices the actor might have made if he had complete knowledge and perfect processing abilities. In reality, of course, searching for and processing information does involve costs, in the form of time, energy, and often money. Most actors either do not want to expend the resources required for comprehensive search and processing or recognize that comprehensive search and processing would not be achievable at any realistic cost. Accordingly, most actors put boundaries on the amount of search they engage in before making decisions. To put it differently, in making decisions actors often consciously choose to be in a state of rational ignorance—“rational” because the expected incremental cost of achieving complete knowledge concerning and deliberating on the choice would be more than the expected incremental gain from making the choice with complete rather than partial knowledge and deliberation. So, for example, a patient may seek a second or perhaps a third doctor’s opinion about treatment, but no more, because he believes it unlikely that consulting an extra five, ten, or fifteen doctors will not improve his choice of treatment enough to justify spending the extra time and money.

Although the concept of bounded rationality is sometimes associated only with limited information and limited information-processing, in fact the concept includes a third element: limited information-processing ability. Even when information is not bounded, the ability to process information is constrained by limitations of computational ability, ability to calculate consequences, ability to organize and utilize memory, and the like. Accordingly, actors will often imperfectly process even the information they do acquire. Thus the principle of bounded rationality “has been defined by [Herbert] Simon as follows: ‘*The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behavior in the real world. . . .*’ [The principle] refers both to neurophysiological limits on the capacity to receive, store, retrieve, and process information without error and to definitional limits inherent in language” (Williamson 1975).

In short, human rationality is normally bounded by limited information, limited information processing, and limited information-processing ability.

## 4.2 Tacit Assumptions

Under rational-actor psychology, actors make choices by considering alternative courses of action and evaluating and comparing their likely outcomes. Often this is true but often it isn't, because actors frequently make choices on the basis of tacit assumptions that the actors do not explicitly consider. The concept of a tacit assumption has been explicated as follows by Lon Fuller:

Words like “intention,” “assumption,” “expectation” and “understanding” all seem to imply a conscious state involving an awareness of alternatives and a deliberate choice among them. It is, however, plain that there is a psychological state that can be described as a “tacit assumption,” which does not involve a consciousness of alternatives. The absent-minded professor stepping from his office into the hall as he reads a book “assumes” that the floor of the hall will be there to receive him. His conduct is conditioned and directed by this assumption, even though the possibility that the floor has been removed does not “occur” to him, that is, is not present in his conscious mental processes. (Fuller and Eisenberg 2006, 732–33).

A more colloquial expression that captures the concept of tacit assumptions is “taken for granted.” As this expression indicates, tacit assumptions are as real as explicit assumptions. Tacit assumptions are not made explicit, even where they are the basis of a choice, precisely because they are taken for granted. They are so deeply embedded that it simply doesn't occur to the parties to make them explicit—any more than it occurs to Fuller's professor to think to himself, every time he is about to walk through a door, “Remember to check my assumption that the floor is still in place.”

Tacit assumptions are closely related to bounded rationality. If actors had infinite time to think about proposed contracts, and no costs for making all of their tacit assumptions explicit, then they might engage in extensive introspectives to determine their tacit assumptions, make them explicit, and put them in the mix of information. But actors do not have infinite time, and introspection is not cost free. It would be irrational to take the time and incur the cost to determine and make explicit every tacit assumption, because the time and cost of doing so would often approach or exceed the expected benefit of the contract. Moreover, normally it would be impossible to make such a determination. As Randy Barnett points out:

[When we add] to the infinity of knowledge about the present world the inherent uncertainty of future events. . . we immediately can see that the seductive idea that a contract can. . . articulate every contingency that might arise before. . . performance is sheer fantasy. For this reason, contracts must be silent on an untold number of items. And many of these silent assumptions that underlie every agreement are as basic as the assumption that the sun will rise tomorrow. They are simply too basic to merit mention. (Barnett 1999, 1163)

In short, in contracting, as in other parts of life, some things go without saying. And a central characteristic of things that go without saying is—they aren't said. As a result,

many steps in an actor's decision-making process are not explicit, and indeed not even conscious.

## 5 THE SECOND WAVE: COGNITIVE PSYCHOLOGY

---

The fundamental premise of rational-actor psychology and therefore Chicago law-and-economics is that contracting actors make decisions rationally. For the most part, this premise is taken as axiomatic. Bounded rationality and tacit assumptions are not *logically* inconsistent with this premise, but when account is taken of those phenomena we get a picture of decision-making that is dissonant with the simple picture of informed choice and explicit reasoning that forms the substratum of rational-choice theory. More important, beginning around the 1970s, under the leadership of Daniel Kahneman and Amos Tversky a large body of work emerged, based on experience, experiment, and theory, which showed that rational-choice theory is often incorrect, lacks explanatory power, or both, because it is not based on the actual psychology of choice. As Tversky and Kahneman stated, “‘Rational-choice theory’ emerged from a logical analysis of games of chance rather than from a psychological analysis of risk and value. The theory was conceived as a normative model of an idealized decision maker, not as a description of the behavior of real people” (Tversky and Kahneman 1986, S251). This new body of work first came to be known as cognitive psychology, but later morphed into a still larger body of work known as behavioral psychology. The departures from rationality identified by cognitive psychology for the most part fell into two major categories: defects in disposition and defects in capability.

### 5.1 Defects in Disposition

Although bounded rationality does not necessarily lead actors to make irrational decisions, two bodies of empirical evidence show that under certain circumstances, actors are often systematically irrational; that is, they often fail to make rational decisions even within the bounds of the information they have acquired.

One body of evidence concerns disposition. For example, people are unrealistically optimistic as a systematic matter. Nearly 90 percent of drivers believe they drive better than average (Weinstein 1980). Ninety-seven percent of consumers believe that they are either average or above average in their ability to avoid accidents involving bicycles and power mowers. In a study by Viscusi and Magat, in which consumers were informed of the true average risks presented by bleach and drain cleaner, only 3 percent of consumers considered their homes to present an above-average risk of hand burn and child poisoning from the use of drain cleaner or gas poisoning or injury to children from the

use of bleach. Roughly half considered their homes to be about average in risk and the other half believed their household had lower-than-average risk. Consumers were particularly optimistic about child poisoning from drain cleaner, which was in fact by far the most severe risk. Sixty-five percent considered their family's risk from this hazard to be below average, while only 3 percent considered their family's risk from this hazard to be above average (Viscusi and Magat 1987). Similarly, Baker and Emery asked subjects who were about to get married to report on their own divorce-related prospects as compared to the divorce-related prospects of the general population. The disparities between the subjects' perceptions of their own prospects and those of the general population were enormous, and were almost invariably in the direction of optimism. For example, the subjects estimated that 50 percent of American couples will eventually divorce, while their own chance of divorce was zero. Similarly, the subjects' median estimate of what percentage of spouses pay all court-ordered alimony was 40 percent, but predicted that their own spouse would pay all court-ordered alimony (Baker and Emery 1993). See generally the chapter 13 by Williams in this volume.

Another deviation from economic rationality involving disposition involves loss aversion. From the perspective of a rational actor, avoiding a loss of \$1,000 is equivalent to forgoing a gain of \$1,000. However, this is not the perception of real actors. Instead, real actors systematically have a much stronger preference for avoiding a loss of \$1,000 than for acquiring a gain of \$1,000. To put this more generally, actors generally feel that the disutility generated by a loss is greater than the utility produced by an identical gain. (Zamir 2012).

## 5.2 Defects in Capability

In making decisions, actors frequently employ heuristics, that is, shortcuts that enable actors to make decisions on the basis of less-than-full information. The use of heuristics goes hand in hand with bounded rationality, because they enable actors to make a decision without full information. Therefore, the use of a heuristic is rational, provided that the heuristic is rational. Many, perhaps most, heuristics are rational. However, a defect in capability is present where actors systematically employ heuristics that are irrational or otherwise faulty. Cognitive psychology has shown that actors commonly do just that—that is, actors commonly display defects in capability by systematically using faulty heuristics. As Tversky and Kahneman conclude, “[T]he deviations of actual behavior from the normative model are too widespread to be ignored, too systematic to be dismissed as random error, and too fundamental to be accommodated by relaxing the normative system” (Tversky and Kahneman 1986, S252).

For example, actors tend to make decisions on the basis of data that is readily available to their memory, rather than on the basis of all the relevant data. This is known as the availability heuristic (Dawes 1988, 10–14, 146–63; Tversky and Kahneman 1982, 164, 166, 174–75). As a result, actors systematically give undue weight to instantiated evidence as compared to general statements, to vivid evidence as compared to pallid

evidence, and to concrete evidence as compared to abstract evidence. Actors are also systematically insensitive to sample size and therefore erroneously take small samples as representative (Arrow 1982, 5; Nisbett and Ross 1980).

Another defect in capability concerns the ability of actors to make rational comparisons between present and future states of the world. For example, the sample consisting of present events is often wrongly taken to be representative, and therefore predictive, of future events. Actors also systematically give too little weight to future benefits and costs as compared to present benefits and costs. Thus Martin Feldstein concludes that “some or all individuals have, in Pigou’s . . . words, a ‘faulty telescopic faculty’ that causes them to give too little weight to the utility of future consumption” (Feldstein 1985, 307).

A defect of capability related to faulty telescopic faculties is the systematic underestimation of risks. Based on the work of cognitive psychologists, Kenneth Arrow observes that “[i]t is a plausible hypothesis that individuals are unable to recognize that there will be many surprises in the future; in short, as much other evidence tends to confirm, there is a tendency to underestimate uncertainties” (Arrow 1982, 9). In fact, empirical evidence shows that actors often not only underestimate but ignore low-probability risks (Arrow 1982; Kunreuther and Slovic 1978).

Still another defect in capability is known as framing. A basic assumption of expected-utility theory, sometimes called invariance, is that a decision-maker’s preference between two options should not depend on how a choice is framed, that is, characterized and presented. Invariance requires that actors make choices based on real consequences, so that two characterizations or presentations of the same option should lead to the same choice. It has been repeatedly shown, however, that whether substantively identical options are framed as gains or losses has a decisive effect on actors’ choices. In a famous experiment, Tversky and Kahneman presented subjects with two problems concerning a choice of alternative programs to combat a disease that would otherwise kill 600 people (Tversky and Kahneman 1981). In problem I, the subjects were told that program A would save 200 lives, while program B carried a one-third probability of saving 600 lives. In problem II, the subjects were told that program A would cost 400 lives, while program B carried a two-thirds probability of losing 600 lives. A moment’s reflection shows that each program A is identical, as is each program B. Nevertheless, because problem I was framed in terms of gains (lives saved) and therefore invoked risk-aversion, while problem II was framed in terms of losses (lives lost) and therefore invoked risk-preference, 72 percent of the respondents chose program A in problem I, while 78 percent chose program B in problem II. Similarly, actors place a higher negative value on losses than on large gains. If an actor is faced with a choice between two economically identical transactions, and in one transaction the choice is framed as avoiding a loss while in the other the choice is framed as foregoing a gain, most actors will irrationally make different choices depending purely on which way the choice is framed (Tversky and Kahneman 1986; Zamir 2012).

These systematic defects in capability are illustrative, not exclusive.

## 6 THE EFFECT OF BOUNDED RATIONALITY, TACIT ASSUMPTIONS, DEFECTS IN COGNITION, AND DEFECTS IN DISPOSITION ON SELECTED AREAS OF AMERICAN CONTRACT LAW

---

Unlike Chicago law-and-economics, the first two waves of behavioral economics scholarship did not directly address the content of legal rules. Once behavioral economics was in place, however, its application to law, including contract law, came easily. Section 6 will discuss the application of behavioral psychology to five areas of American contract law: liquidated damages, express conditions, mutual mistake, unexpected circumstances, and form contracts.

### 6.1 Liquidated Damages

It is a basic principle of contract law that absent fraud, duress, unconscionability, or the like, bargain contracts (“contracts”) will be enforced according to their terms, that is, without judicial review of the desirability of the terms. One exception to this principle concerns liquidated damages, that is, damages that are fixed in the contract. Although the details of the law of liquidated damages are complex, the basic principle is that, unlike most other contractual provisions, liquidated-damages provisions are given special scrutiny by the courts. Many commentators have criticized this principle. Most of the critiques have an implicit or explicit three-part structure: (1) They begin by assuming that the major justification for the principle of special scrutiny is that liquidated-damages provisions lend themselves to blameworthy exploitation and consequent one-sidedness in a way that other types of contract provisions do not. (2) It is then argued that this justification will not hold. (3) The conclusion is that the special-scrutiny principle is therefore unjustified. So, for example, Goetz and Scott argue that the special scrutiny of liquidated-damages provisions arose in a historical context in which protections against fraud and duress were not available. Given the modern development of unconscionability as a unifying unfairness principle, they suggest, the law should simply collapse the treatment of liquidated damages into that principle (Goetz and Scott 1977, 592).

In fact, however, the primary justification for the special scrutiny of liquidated-damages provisions is not that such provisions are uniquely amenable to blameworthy exploitation and one-sidedness, but that such provisions are systematically likely to reflect the limits of cognition, and in particular, bounded rationality, defects in disposition, and defects in capability.

Begin with bounded rationality. Contracting parties normally will find it relatively easy to evaluate proposed *performance terms*—terms that specify what performance each party is required to render, such as subject matter, quantity, and price. In contrast, at the time a contract is made it is often impracticable if not impossible to imagine all the possible scenarios of *breach* and, therefore, all the ways in which a liquidated-damages provision might operate. Similarly, the inherent complexity of determining the application of a liquidated-damages provision to every possible scenario of breach will often exceed actors' information-processing abilities.

Even on the doubtful assumption that a contracting party could imagine all the possible scenarios of breach and determine the application of a liquidated-damages provision to each scenario, the benefits of such extensive search and information-processing will often appear to be much less than the costs. A party who contracts to buy or sell a commodity normally expects to perform. Accordingly, the expected benefit of carefully deliberating on performance terms is compelling and the expected cost of such deliberation usually does not outweigh the expected benefit. In contrast, parties may not expect that a liquidated-damages provision will ever come into play, partly because they optimistically intend to perform and partly because experience will tell them that in general there is a high rate of performance of contracts. For example, if contracts are substantially performed at least 95 percent of the time (which observation suggests is likely), all the costs of processing the more remote applications of a liquidated-damages provision would have to be taken into account, but the benefits of such processing would have to be discounted by 95 percent. The resulting cost-benefit ratio will often provide a substantial disincentive for processing every possible application of a liquidated-damages provision, even if it were possible to imagine every such application. As a result, contracting parties are often unlikely to completely think through liquidated-damages provisions, and therefore are often unlikely to fully understand the full implications of such provisions.

Defects in disposition also play a role. Because actors tend to be unrealistically optimistic, a contracting party will probably believe that her performance is more likely, and her breach is less likely, than is actually the case. Accordingly, unjustified optimism will reduce even further the deliberation that actors give to liquidated-damages provisions.

Finally, defects in capability have special relevance to liquidated-damages provisions. Because actors tend to take the sample of present evidence as unduly representative of the future, a contracting party is likely to overestimate the extent to which his present intention to perform is a reliable predictor of his future intention. Because actors have faulty telescopic faculties, a contracting party is likely to overvalue the benefit of performance, which will normally begin to occur in the short term, as against the cost of breach, which will typically occur only down the road. Because actors tend to underestimate risks, a contracting party is likely to underestimate the risk that a liquidated-damages provision will take effect against her. And the availability heuristic may lead a contracting party to give undue weight to her present intention to perform,



which is vivid and concrete, as compared with the abstract possibility that future circumstances may lead him to breach.

The limits of cognition not only justify the liquidated-damages principle but affect the point of time at which special scrutiny should be focused. If the justification for giving special scrutiny to liquidated-damages provisions is that such provisions are especially subject to blameworthy exploitation and one-sidedness, special scrutiny should be focused on the time the contract is made—an ex ante, forward-looking test. In contrast, if the justification for special scrutiny is that parties often make cognitive errors in adopting such provisions, special scrutiny should be focused on the time of breach—an ex post, second-look test, which compares the amount of liquidated damages with the promisee's actual loss—on the ground that a gross discrepancy between forecast and result suggests that the liquidated-damages provision was a product of limited or defective cognition (Eisenberg 1995).

## 6.2 Express Conditions

Just as contract law has a special principle governing liquidated-damages provisions, so too does it have a special principle governing express conditions. A promise is a commitment that some specified state of events will or will not occur. An express condition qualifies a promissory duty, by providing either that a party is not obliged to perform the duty unless a specified event occurs or fails to occur or that the duty will be suspended or terminated if a specified event occurs or fails to occur.

Promises and express conditions have much different consequences. If a party has substantially performed her promises she normally can enforce her contract even though she has not perfectly performed, subject to damages for the imperfect performance. In contrast, the nominal rule that governs express conditions is that a party cannot enforce a contract unless the conditions to her counterparty's obligation to perform have been perfectly fulfilled. Similarly, a promisee normally cannot terminate a contract on the ground of the promisor's breach unless the breach is material. In contrast, if one contracting party, A, fails to perfectly fulfill a condition to the performance of the other party, B, the nominal rule is that B can terminate the contract even if A's failure is immaterial. Termination for nonfulfillment of an express condition is a very severe sanction, because A may lose not only the value of the contract but also the value of her prior performance, subject to possible recovery for any benefit conferred.

The nominal rule that express conditions must be perfectly fulfilled is qualified in several ways. Most notable is an ant forfeiture principle, which is a variant of the special-scrutiny principle concerning liquidated damages. The ant forfeiture principle is codified as follows in *Restatement Second of Contracts* Section 229: "To the extent that the non-occurrence of a condition would cause disproportionate forfeiture, a court may excuse the non-occurrence of that condition unless its occurrence was a material part of the agreed exchange." The comment to Section 229 defines forfeiture as "the denial of compensation that results when the obligee loses his right to the agreed

exchange after he has relied substantially, as by preparation or performance on the expectation of that exchange.” The comment continues:

In determining whether the forfeiture is disproportionate, a court must weigh the extent of the forfeiture by the obligee against the importance to the obligor of the risk from which he sought to be protected and the degree to which that protection will be lost if the non-occurrence of the condition is excused to the extent required to prevent forfeiture.

The comment also makes clear that unlike a determination of unconscionability, application of the forfeiture principle depends on the actual result of enforcing the express condition, rather than on the conduct of the parties at the time the contract is formed. Thus, the forfeiture principle is explicitly based on an *ex post*, second-look approach.

The cases amply support the antiforfeiture principle. For example, in *Holiday Inns of America, Inc. v. Knight* (450 P.2d 42 (1969)), a buyer, B, entered into an option contract with a seller, S, for the purchase of property. The contract provided for an initial payment of \$10,000 by B and four additional \$10,000 payments due on July 1 of each successive year. Under the contract, failure to make a payment on or before the specified date each year would result in cancellation of the option. B made timely payments for two years, and also made expenditures that substantially increased the value of the optioned property. In the third year, however, S received B’s check on July 2, one day late. S returned the check, stating that the option was canceled because the payment condition had not been fulfilled. The court refused to enforce the cancellation provision on the ground that it would work a forfeiture.

The principle that governs the enforcement of express conditions is very similar to the principle that governs the review of liquidated-damages provisions. Both principles concern sanctions. Both principles allow the courts to override bargained-for provisions even in the absence of unconscionability. Both principles turn on an *ex post*, second-look approach. Most importantly, the principle that governs the excuse of express conditions, like the principle that requires special scrutiny of liquidated-damages provisions, is best explained by the limits of cognition. Were it not for the limits of cognition, the law should no sooner excuse the imperfect fulfillment of an express condition than it should refuse to enforce performance terms that turn out to be extremely disadvantageous to one party. However, the limits of cognition operate with respect to express conditions in a manner that closely parallels the operation of those limits with respect to liquidated-damages provisions.

To begin with, bounded rationality will often limit a party’s full comprehension of express conditions. The costs of determining the various ways an express condition may fail to be perfectly fulfilled can be very high, because often at the time of contracting a party cannot efficiently conceive of every contingency under which nonfulfillment may occur. Indeed, not every contingency may even be practicably conceivable at the time of contracting. Moreover, the benefits of incurring these costs are likely to seem dubious to a party who is on the verge of making a contract. Just as a liquidated-damages provision bites only if a party fails to perform a promise, so an

express condition bites only if it is not fulfilled. Because parties normally expect to fulfill conditions, at the time a contract is made the consequences of nonfulfillment will seem remote. Accordingly, a party is likely to view the cost of fully deliberating on the operation of an express condition as unduly high in light of the significant probability that the provision will never come into play. Although parties may give more deliberation to an express condition than to a liquidated-damages provision because the operation of an express condition is apt to be more specific, parties may also rationally (but mistakenly) give express conditions less deliberation because they may be unaware of the potentially draconian legal sanctions for immaterial variations from perfect fulfillment. Indeed, most parties are probably not aware of either the legal distinction between promises and express conditions or of the consequences of that distinction, which often present difficult questions even for courts.

Defects in disposition also have a special impact on express conditions. Because actors are unduly optimistic, a contracting party is likely to believe that fulfillment of an express condition is more likely, and nonfulfillment less likely, than is actually the case. Similarly, if a contracting party realized that his imperfect fulfillment of a condition, however slight, would involve an extremely large loss that was wholly disproportionate to the significance of the nonfulfillment, because of loss-aversion it is unlikely that he would agree to a perfect-fulfillment rule. Finally, like the subjects in Baker and Emery's (1993) study, contracting parties may have an unduly optimistic belief that their partners will act fairly and refrain from invoking harsh sections.

Next, defects in capability are likely to discourage parties from focusing on express conditions. The availability heuristic is likely to lead a contracting party to give undue weight to his present intention to fulfill an express condition, which is vivid and concrete, as compared to the possibility that future circumstances may lead to nonfulfillment, which is pallid and abstract. Moreover, the tendency to underestimate risks will likely cause a contracting party to underestimate the risk that an express condition will not be fulfilled.

It might be argued that even if one party to a contract, A, would be reluctant to agree to an express condition if he understood that if the condition was not perfectly fulfilled he would face a draconian sanction for immaterial imperfections, B would insist on that sanction. That is possible, but unlikely. If both parties fully understand the operation of the condition, then the price that B must pay for A's performance will be higher than it otherwise would be, to reflect A's additional risks. Given perfect knowledge by both parties, B would probably prefer to pay less, without the power to impose draconian sanctions for immaterial nonfulfillment of a condition, than to pay more, with that power.

Because the principle governing the excuse of express conditions to prevent forfeiture is best justified by the limits of cognition, that principle should be interpreted in a manner parallel to the interpretation of the liquidated-damages principle: If, in the scenario of imperfect fulfillment that actually occurred, a requirement of perfect fulfillment would result in a substantial loss to one party that is significantly out of proportion to the interest of the other in perfect fulfillment, then courts should not

require perfect fulfillment unless it is established that the parties had a specific and well-thought-through intention that perfect fulfillment should be required in a scenario like the one that actually occurred.

### 6.3 Mutual Mistake

Suppose A and B enter into a contract that is based on a material mistaken assumption, held by both parties, about the outside world. This kind of mistake is known as a mutual mistake (or, under the law of England and some other countries, a “common mistake”), and often serves to make the contract unenforceable, if it has not been performed, or reversible, if it has been. The difficult question is when and why such a mistake should be a basis for relief. The answer lies partly in the limits of cognition, and more particularly in the role of tacit assumptions.

Assume first that a shared factual assumption is made explicit in a contract. In that case, if the assumption turns out to have been mistaken, normally it should furnish a basis for relief as a matter of interpretation. This can be illustrated by a hypothetical variations of a leading case, *Griffith v. Brymer* (19 T.L.R. 434 (K.B. 1903)). Edward VII was to be crowned in Westminster Abbey on June 26, 1902, following a coronation procession from Buckingham Palace to the Abbey (Wladis 1987, 1609–10). Brymer had a room that overlooked the proposed route of the procession. On June 24, Griffith entered into an oral agreement to take the room for the purpose of viewing the procession, at the price of £100, and delivered a check for that amount. Unknown to either party, that morning Edward’s physicians had decided that Edward required surgery, and as a result the coronation and procession had been postponed. Griffith sued to recover the £100. Now assume that the contract had explicitly stated, “This agreement is made on the assumption that the coronation and procession are still on as of this moment.” It could then be concluded that if the assumption was incorrect, Griffith would recover his payment under a relatively straightforward interpretation of the contract.

Against this background, the principle that governs shared mistaken tacit assumptions should be, and implicitly is, as follow. A shared mistaken tacit assumption should give rise to relief where the assumption would give rise to relief if it had been explicit rather than tacit, because a tacit assumption is as real as an explicit assumption, and a contracting party should not be penalized for acting in the way most contracting parties do, that is, not making every tacit assumption explicit.

### 6.4 Unexpected Circumstances

The doctrine of unexpected circumstances, which includes impossibility, impracticability, and frustration, strongly resembles the doctrine of mutual mistake, because it is based in significant part on the role of mistaken shared tacit assumptions—although the tacit assumptions that figure in mutual-mistake cases concern a present state of

the world while the tacit assumptions that figure in unexpected-circumstances cases concern an expected state of the world. Consider *Krell v. Henry* (2 K.B. 740 (1903)), another famous coronation case. Krell, who had a flat in London, had left England in March 1902. Before he left, Krell instructed his solicitor to let the flat on such terms, and for such period, not to exceed six months, as the solicitor thought proper. On June 17, Henry noticed an announcement on the exterior of Krell's flat that windows to overlook the coronation procession were to be let. Henry then entered into an agreement, confirmed in writing on June 20, to take the flat for the days, but not the nights, of June 26 and 27, for £75, and put down £25 as a deposit. After the coronation and procession were postponed, Henry did not pay the remaining £50, and Krell sued for that amount. The court held for Henry. In the leading opinion, Lord Justice Vaughan Williams said, "I think it cannot reasonably be supposed to have been in the contemplation of the contracting parties when the contract was made, that the coronation would not be held on the proclaimed days. . . or along the proclaimed route" (id., p. 750; Wladis 1987, 1618).

Now suppose the agreement in *Krell v. Henry* had explicitly stated, "This agreement is made on the assumption that the coronation procession will take place in six days as scheduled." In that case, an analysis based on unexpected circumstances would have been unnecessary. Instead, Henry would have prevailed as a matter of interpretation. In the actual case the relevant assumption was tacit, rather than explicit. That should make no difference, for the same reasons it should make no difference in mutual-mistake cases.

Loss-aversion also plays a special role in unexpected-circumstances cases. In mutual-mistake cases the plaintiff is usually a buyer who finds that because of a mistake the value of the commodity he contracted to purchase is not what he reasonably expected it to be. In such cases, the buyer stands to lose all of the contract price, but no more. To put it differently, in such cases the buyer's loss, while it may be severe, is bounded. In contrast, in unexpected-circumstances cases the plaintiff is usually a *seller* who finds that due to the occurrence of the unexpected circumstances his cost for performing the contract is significantly higher than the contract price. In some cases the increase is bounded, but in other cases it is not. The reason is that where the seller's cost of performance dramatically and unexpectedly rises above the contract price the increase will often be market-wide. In that event, the increase normally will increase not only the seller's cost but also the market value of the contracted-for commodity. Therefore, in the absence of judicial relief, the buyer's expectation damages, based on the difference between the contract price and the market price, can rise to a very high level that is not bounded by the contract price. In such cases, the seller's loss is one that he almost certainly would not have agreed to *ex ante*.

Of course, this will not be true if the parties recognize that their assumption is problematic, if the risk is explicitly or implicitly allocated to the adversely affected party, if the assumption is not shared or is evaluative, or if the impact of the unexpected circumstances is not material. In many cases, however, none of those exceptions is likely to apply. For example, in *Krell v. Henry* we can be pretty confident that (1) actors in the positions of the contracting parties would have shared the tacit assumption that

the coronation would take place in six days, (2) the contract was made on the basis of that assumption, (3) Henry was not assuming the risk that the assumption was incorrect—was not gambling, and was not being paid to gamble, on whether the coronation would take place, and (4) due to the phenomenon of loss-aversion, the impact on Henry of a £150 out-of-pocket loss would be greater than the impact on Krell of a £150 forgone anticipated gain.

## 6.5 Form Contracts

The problems raised by the use of form contracts have been a major preoccupation of contract-law scholars for many years. The primary areas of concern have been the enforceability of pre-prepared terms and the import of pre-prepared terms in determining whether a form sent in response to an offer constitutes an acceptance. Both concerns rest ultimately on the limits of cognition.

To begin with, most pre-prepared terms are nonperformance terms that relate to the future and concern low-probability risks. Accordingly, the cognitive problems associated with liquidated-damages provisions and express conditions, including bounded rationality, unduly optimistic disposition, systematic underestimation of risks, and undue weight on the present as compared with the future, apply as well to most pre-prepared terms.

Of these, the phenomenon of rational ignorance plays a particularly powerful role. Call a party who prepares a form contract a form-giver, and a party who receives a form contract a form-taker. Typically, a form-giver offers a package consisting of a physical commodity and a form contract that states the terms on which the physical commodity is sold. Each part of the package, in turn, consists of a number of subparts: The commodity has physical attributes, such as size, shape, and color; the form contract has business attributes, such as price and quantity; and legal attributes—essentially, non-performance terms—such as limitations on remedies.

To make an optimum substantive decision, the form-taker would carefully deliberate on all of the form's legal attributes. Analyzing these attributes, however, would often be unduly costly. First, a form contract often contains a very large number of legal terms. Form insurance contracts, for example, typically include thirty, forty, or more legal terms. Moreover, the meaning and effect of the pre-prepared provisions will very often be inaccessible to laypersons. In part, this is because the terms are often written in exceedingly technical prose. Even if the terms are written clearly, however, the form-taker usually will be unable fully to understand their effects, because pre-prepared terms characteristically vary the form-taker's baseline legal rights, and most consumers do not know their baseline rights.

*Williams v. Walker-Thomas Furniture Co.* (350 F.2d 445 (D.C. Cir. 1965)) is a good example of both problems. A consumer regularly purchased furniture and home appliances from a seller on installment credit. Each purchase agreement contained the following preprinted provision:



[T]he amount of each periodical instalment payment to be made by [purchaser] to the Company under this present lease shall be inclusive of and not in addition to the amount of each installment payment to be made by [purchaser} under. . . prior leases, bills or accounts; and all payments now and hereafter made by {purchaser} shall be credited pro rata on all outstanding leases, bills and accounts due the Company by [purchaser} at the time each such payment is made.

The immediate but obscure effect of this provision was that the seller retained title to each item purchased under *any* contract until the buyer finished paying in full for all items purchased under *all* contracts, even though the balance due on any particular item might be worked down to a few cents, as indeed it was in *Williams* (Skilton and Helstad 1967).

The effects of this provision on the buyer's baseline legal rights were even more obscure and undoubtedly were unknown to the buyer. Under this provision, until the buyer brought her total unpaid balance to zero, the seller could repossess every item purchased under every contract made while a balance was still outstanding, through summary process under the replevin statute, thereby circumventing the normal process of a full trial. Moreover, by virtue of this provision, no item the buyer purchased under any contract with the seller made while a balance was outstanding would fall within the protection of statutes exempting defined classes of property from attachment to satisfy a judgment.

Despite the significance of this provision, few laymen would have understood all or even any of its implications. Even legal experts often can't understand the pre-prepared items of form contracts. During the oral argument of *Gerhardt v. Continental Insurance Cos.* (224 A.2d 328 (1966)) before the great New Jersey Supreme Court of the time, Chief Justice Weintraub looked at the insurance policy at issue and said, "I don't know what it means. I am stumped. They say one thing in big type and in small type they take it away." Justice Haneman added, "I can't understand half of my insurance policies." Justice Francis said, "I get the impression that insurance companies keep the language of their policies deliberately obscure" (New Jersey Court Overrules Small Print in Policy 1969).

The bottom line is simple: The verbal and legal obscurity of pre-prepared terms renders the cost of reading and deliberating on these terms exceptionally high. In contrast, the low probability of these terms coming into play heavily discounts the benefit of reading and deliberation. Furthermore, the length and complexity of form contracts is often not correlated to the dollar value of the transaction. Where form contracts involve a low dollar value, the cost of thorough reading and deliberation, let alone the cost of seeking legal advice about their meaning and effect, will usually be prohibitive in relation to the benefit. Faced with pre-prepared terms whose effect the form-taker knows he will find difficult or impossible to fully understand, which involve risks that probably will never mature, which are unlikely to be worth the cost of reading and processing, and which probably aren't subject to revision in any event, a rational form-taker will typically decide to remain ignorant of the terms (Meyerson 1990).<sup>1</sup>

<sup>1</sup> For more detailed behavioral analysis of standard-form contracts see Becher (2007).



## 7 THE THIRD WAVE: THE BEHAVIOR OF CONTRACTING PARTIES

---

The first and second waves of behavioral psychology consist of scholarship that does not address contract law as such but does provide reasons why the *content* of given contract-law rules should be a certain way. In contrast, the third wave bears principally on how contracting actors *behave*—for example, how contracting actors think about sanctions for breach and are incentivized or disincentivized by given types of contractual provisions. Some third-wave scholarship, however, indirectly concerns the content of contract law—for example, by rebutting central tenets of Chicago law-and-economics—and some third-wave scholarship bears more directly on the content of contract law. In what follows I will survey some of the central contributions to this body of scholarship.<sup>2</sup>

### 7.1. The Option Theory of Contracts and the Theory of Efficient Breach

Two of the central tenets of Chicago law-and-economics are the option theory of contracts and the theory of efficient breach.

The option theory of contracts is summarized as follows by Joseph Perillo:

Oliver Wendell Holmes’s most notorious statement about contract law was that “[t]he duty to keep a contract at common law means a prediction that you must pay damages if you do not keep it—and nothing else.” This generally has been interpreted to mean that a contracting party has a lawful option to perform or not. (Perillo 2000)

(Perillo argues forcefully that Holmes did not in fact hold to the option theory, but that is irrelevant for present purposes, since Chicago law-and-economics scholars have adopted the theory and believe it stems from Holmes.) The theory of efficient breach, which builds on the option theory of contracts, holds that breach of contract is efficient, and therefore desirable, if the promisor’s profit from breach, after payment of expectation damages, will exceed the promisee’s loss from breach. Probably the best-known exposition of the theory is that given by Richard Posner in his book *Economic Analysis of Law*. This exposition has changed somewhat over the eight editions of that book. Here is the core of the exposition in the first edition:

<sup>2</sup> Additional noteworthy contributions include Eigen (2012a, 2012b), Feldman, Schurr, and Teichman (2014), Korobkin (1998a, 1998b), Sunder (2012), Teichman (2010), Vanberg (2008), and Wilkinson-Ryan (2010, 2011, 2012).

[I]n some cases a party [to a contract] would be tempted to breach the contract simply because his profit from breach would exceed his expected profit from completion of the contract. If his profit from breach would also exceed the expected profit to the other party from completion of the contract, and if damages are limited to loss of expected profit, there will be an incentive to commit a breach. There should be. (Posner 1972, 57)

Although the theory of efficient breach is typically presented in very generalized terms, the theory can only be properly understood and evaluated in the context of paradigm cases to which it might be meaningfully applied. The most salient of these is the resale paradigm. In this paradigm, a seller who has contracted to sell a commodity to a buyer breaches the contract in order to sell the commodity to a third party who comes along later and offers a higher price. For example, here is the centerpiece illustration of the theory of efficient breach in the first edition of *Economic Analysis of Law*:

I sign a contract to deliver 100,000 custom-ground widgets at \$.10 apiece to A, for use in his boiler factory. After I have delivered 10,000, B comes to me, explains that he desperately needs 25,000 custom-ground widgets at once since otherwise he will be forced to close his pianola factory at great cost, and offers me \$.15 apiece for 25,000 widgets. I sell him the widgets and as a result do not complete timely delivery to A, who sustains \$1000 in damages from my breach. Having obtained an additional profit of \$1250 on the sale to B, I am better off even after reimbursing A for his loss. Society is also better off. Since B was willing to pay me \$.15 per widget, it must mean that each widget was worth at least \$.15 to him. But it was worth only \$.14 to A—\$.10, what he paid, plus \$.04 (\$1000 divided by 25,000), his expected profit. Thus the breach resulted in a transfer of the 26,000 widgets from a lower valued to a higher valued use. (Posner 1972, 72)

The theory of efficient breach is defective in a number of respects. For example, the theory is based on the premise that the expectation measure makes a promisee indifferent between performance and damages, but that is not the case. Among other things, expectation damages are based on the objective value of a promised performance rather than on the promisee's subjective value; the promisee's damages for its losses from breach are often cut off by various rules, such as the principle of *Hadley v. Baxendale*; and under US law expectation damages normally do not include legal fees and other costs of litigation.

In addition, both the option theory of contracts and the theory of efficient breach run counter to experience and intuition. Take this simple thought experiment: Suppose a seller and a buyer have negotiated an agreement under which the seller agrees to sell a differentiated commodity to the buyer—say a home to live in or a used machine that the buyer will employ as a factor of production. As the parties are about to sign a written contract, the seller says, "In all honesty, I should tell you that although I have no present intention to breach this contract, neither do I have a present intention to perform. If a better offer comes along I will take it and pay you expectation damages. In fact, I will begin actively looking for a better offer right after we sign this contract. Let's insert a provision that recognizes that I will do just that." What would be the buyer's likely response? Under the option theory of contracts and the theory of efficient breach,

the buyer would say, “Of course, I expect no more.” Experience and intuition tell us, however, that most buyers would be surprised if not shocked by such a statement, and would either walk away, insist on an explicit contractual provision stating that the seller has a present intent to perform and that any profit on breach and resale will go to buyer, or demand a payment in the form of a reduced price for the seller’s right to resell. Buyers will react this way because normally one of the very points of a bargain promise is to convince the promisee that the promisor will perform.

Third-wave scholarship provides experimental evidence that supplements these experiential and intuitive understandings. For example, in an experiment conducted by Daphna Lewinsohn-Zamir (with both laypersons and experienced businesspeople as subjects) involving a breach of contract, 69 percent of the subjects preferred in-kind relief and only 26 percent chose monetary relief. More tellingly,

[A] Questionnaire. . . depicted a commercial transaction for the purchase of iron and presented the buyer with a choice between two contracts. One contract entitled the buyer to delivery of the iron in-kind; the other entitled her to either delivery in-kind or full expectation damages, and emphasized that the monetary award would include not only restitution of the advance payment and compensation for any increase in market price, but also compensation for the delay in supplying the iron and the inconvenience of purchasing substitute iron.

. . . 74% opted for the contract that required the seller to deliver the iron in-kind and 66% of those indicated that they would not agree to sign the other contract, even for a discount in the price of the iron. These results were highly statistically significant. It seems that people prefer actual performance of contracts even when the promised asset is not unique and its value is easily quantifiable. This preference was often powerful enough to result in an outright rejection of the second contract, even with a discount. Although 34% of those initially choosing the in-kind option were willing to switch to the monetary option for a discount in price, 30% of them conditioned the switch on an unrealistically high discount, between 20% and 50% of the contract price. It is extremely doubtful that a seller would be willing to offer a reduction of this magnitude (in addition to full expectation damages) to secure the right to sell the iron to another buyer, in the unlikely event that such a buyer would appear. (Lewinsohn-Zamir 2013, 178–79)

## 7.2 Moral Considerations

Another area in which third-wave scholarship rebuts central tenets of Chicago law-and-economics concerns the role of moral considerations in the contracting process. Chicago law-and-economics either explicitly or implicitly rests on the premise that those elements have no role in that process. For example, the option theory of contracts was part of Holmes’s program to exclude moral considerations from contract law. Similarly, in Posner’s widgets hypothetical the widget-seller is motivated only by money, and no account is taken of the seller’s and buyer’s moral view of breach. This is

particularly strange because the rational-actor model assumes that contracting parties aim to maximize their utility, and adherence to moral and other normative factors is part of the utility function of many actors. Thus an actor may perform a contract to provide a commodity at \$X even where a late-coming third party offers more than \$X, partly because the actor takes morality into effect in making decisions. (This is not to say that all actors will do this, or that actors who do this in some cases will do it all the time. It is to say that moral and other normative factors are likely to often figure in perform-or-breach decisions.)

Although role of moral considerations in the contracting process is clear as a matter of experience and intuitions, third-wave scholarship reinforces those by experimental evidence. For example, Wilkinson-Ryan and Baron (2009) found experimentally that promisees were “more punitive”—that is, thought damages should be higher—when the promisor breached to increase her gain than when she breached to avoid a loss, and when the promisor breached rather than negotiating a termination of the contract. “Our results,” they concluded, “suggest that people are quite sensitive to the moral dimensions of contracts” (Wilkinson-Ryan and Baron 2009, 405; see also Wilkinson-Ryan and Hoffman 2010). Feldman and Teichman found experimentally that perform-or-breach decisions were often motivated in part by moral considerations (Feldman and Teichman 2011). Lewinsohn-Zamir concluded, on the basis of experimental evidence, that for contracting actors “outcomes” included not only material gains and losses but also such factors as the attitude of the other party in rendering performance or a settlement and whether an injurious outcome was brought about intentionally or unintentionally (Lewinsohn-Zamir 2012).

### 7.3 Remedies

Under a basic principle of American contract law, the normal remedy for breach of contract is expectation damages. Two other rules concerning contract remedies reflect that principle. Under one rule, specific performance is an extraordinary remedy, which will be granted only if damages would not be adequate to protect the expectation interest of the injured party (*Restatement Second of Contracts* Section 359). Under another, punitive damages will be awarded only when a breach is independently a tort (Section 355). Both the specific-performance and punitive-damages rules have been called into question in modern legal scholarship (Eisenberg 2005), and both rules have been eroding. As a result, specific performance is much more freely granted than the rule would suggest (Laycock 1990).

Similarly, many jurisdictions have adopted tests for punitive damages that are more expansive than that of Section 355. For example, in *Suffolk Sports Center, Inc. v. Belli Construction Corp.* (212 A.D.2d 241 (N.Y. 1995)), a landlord had barricaded the entrances to the tenant’s leased sports facilities during a dispute concerning the lease. In approving punitive damages against the landlord, the court said that punitive damages were

available where the breach evinces a “‘high degree of moral turpitude’ or is ‘actuated by evil and reprehensible motives,’ and demonstrates ‘such wanton dishonesty as to imply a criminal indifference to civil obligations’” (id., p. 246). In *Miller v. Byrne* (916 P.2d 566, 580 (Colo. App. 1995)), the court said that “Punitive damages are available. . . when a plaintiff is able to prove beyond a reasonable doubt that the defendant engaged in ‘willful and wanton’ misconduct. Willful and wanton conduct means conduct purposefully committed which the actor must have realized is dangerous, done heedlessly and recklessly, without regard to consequences, or of the rights and safety of others, particularly of the plaintiff.”

In practice, punitive damages for breach of contract are not that uncommon. Marc Galanter (2001) analyzed data concerning contract litigation including a Bureau of Justice Statistics study of contract litigation in 1992 in the seventy-five largest counties in the United States, which account for about one-third of the population. That study, Galanter reported, found that 12 percent of all winning plaintiffs in jury-tried contracts cases were awarded punitive damages (as compared to only 4 percent of winning plaintiffs in jury-tried tort cases). Even in judge-tried contracts cases, which outnumbered jury-tried cases by two to one, winning plaintiffs in contracts cases were awarded punitive damages 3.6 percent of the time.

Third-wave behavioral scholarship supports both modern approaches. For example, recall that Lewinsohn-Zamir found that 69 percent of the subjects in an experiment involving a breach of contract preferred in-kind relief to monetary relief even when the contract referred to fugible goods (Lewinsohn-Zamir 2013). Similarly, in an experiment by Wilkinson-Ryan and Baron (2009), the subjects believed that where a promisor breached to make a gain, damages should be significantly higher than expectation damages. A similar result was reported by Wilkinson-Ryan and Hoffman (2010).

## 7.4 Critiques

Critiques of scholarship in the first and second waves have generally been unconvincing. (See chapter 4 by Ulen in this volume). In contrast, some and perhaps much third-wave scholarship is subject to three methodological problems.

First, many or most of the third-wave experiments are promisee-centered; that is, subjects in the experimental studies were usually asked questions that explicitly or implicitly put them in the role of a promisee. Therefore, we don’t know how the subjects would have responded if they had been asked instead, what remedy would be efficient and fair? or, If you were the breaching party, what remedy would you believe appropriate?

Second, in many or most experiments the subjects were placed in the roles of individuals who had contracted on their own behalf and therefore would take breach personally. A large amount of contracting activity does indeed involve individuals contracting on their own behalf, such as consumers or the owners of small businesses. But another large amount of contracting activity involves individuals contracting on

behalf of others, such as purchasing agents, salespeople, and corporate managers. With limited exceptions, the third-wave experiments don't show how the experimental subjects would have reacted if they had been placed in representative roles. This omission is particularly striking because a representative who makes a contract is unlikely to be the person who deals with the fallout from breach. Some experimental findings suggest that people distinguish between individuals and organizations in judging contract breaches.

Third, it is very difficult to operationalize the results of many or most third-wave experiments. Many of the experiments involve only easy issues, such as how promisees feel about breach. They do not consider what negative consequences, if any, would result from incorporating those feelings into legal rules. (An exception is Lewinsohn-Zamir 2013.) For example, increasing the penalties for breach might serve the interests of a promisee in a given contractual transaction, but might disserve social interests—and even the interests of promisees as a class—by dampening the willingness of actors to make contracts. Similarly, although experimental subjects routinely prefer the remedy of specific performance, making that remedy routinely available would present a host of problems (Eisenberg 2005).

## 8. CONCLUSION

---

As regards contract law, Chicago law-and-economics is a relatively closed and often unrealistic system that rests on inaccurate predicates, such as the option theory of contracts and the theory of efficient breach, and on premises that purport to be universally applicable but were not, such as rational-actor psychology and the exclusion of morality from contract law. In contrast, behavioral psychology opens up contract law to reality. It explains many rules of contract law that Chicago law-and-economics and rational-actor psychology cannot explain—such as the special scrutiny of liquidated-damages provisions, the forfeiture principle applicable to express conditions, and the treatment of form contracts—and teaches how these and certain other rules should be applied. At the same time, it leaves ample room for analysis and rules based on conventional microeconomics where appropriate, or a combination of conventional microeconomics and behavioral economics.

## REFERENCES

---

- Arrow, K. J. 1982. Risk perception in psychology and economics. *Economic Inquiry* 20(1), 1–9.
- Baker, L. A., and R. E. Emery. 1993. When every relationship is above average. *Law and Human Behavior* 17, 439–49.
- Barnett, R. E. 1999. *Contracts: Cases and doctrine*. 1st ed. Gaithersburg: Aspen Law and Business.
- Becher, S. 2007. Behavioral science and consumer standard form contracts. *Louisiana Law Review* 68, 117–79.

- Dawes, R. 1988. *Rational choice in an uncertain world*. 1st ed. San Diego: Harcourt Brace Jovanovich.
- Eigen, Z. 2012a. Empirical studies of contract. *Annual Review of Law and Social Science* 8, 291–306.
- Eigen, Z. 2012b. When and why individuals obey contracts: Experimental evidence of consent, compliance, promise, and performance. *Journal of Legal Studies* 41(1), 67–93.
- Eisenberg, M. 1995. The limits of cognition and the limits of contract. *Stanford Law Review* 47(2), 211–59.
- Eisenberg, M. 2005. Actual and virtual specific performance, the theory of efficient breach, and the indifference principle in contract law. *California Law Review* 93, 975–1050.
- Feldman, Y., A. Schurr, and D. Teichman. 2014. Reference points and contractual choices: An experimental examination. *Journal of Empirical Legal Studies* 10, 512–14.
- Feldman, Y., and D. Teichman. 2011. Are all contractual obligations created equal? *Georgia Law Review* 100, 5–52.
- Feldstein, M. 1985. The optimal level of social security benefits. *Quarterly Journal of Economics* 100(2), 303–20.
- Fuller, L. L., and M. A. Eisenberg. 2006. *Basic contract law*. 8th ed. St. Paul, MN: West.
- Galanter, M. 2001. Contract in court; or almost everything you may or may not want to know about contract litigation. *Wisconsin Law Review* 2001, 577–627.
- Goetz, C. J., and R. E. Scott. 1977. Liquidated damages, penalties and the just compensation principle: Some notes on an enforcement model and a theory of efficient breach. *Columbia Law Review* 77(4), 55494.
- Holmes, O. W. 1920. Law in science and science in law. *Collected Legal Papers*. New York: Harcourt, Brace.
- Korobkin, R. 1998a. The status quo bias and contract default rules. *Cornell Law Review* 83, 609–87.
- Korobkin, R. 1998b. The legal implications of psychology: Human behavior, behavioral economics and the law. *Vanderbilt Law Review* 51, 1584–636.
- Kunreuther, H., and P. Slovic. 1978. Economics, psychology, and protective behavior. *American Economic Review* 64, 66–67.
- Langdell, C. C. 1880. *A summary of the law of contracts*. 2nd ed. Boston: Little, Brown and Co.
- Laycock, D. 1990. The death of the irreparable injury rule. *Harvard Law Review* 103, 687–771.
- Lewinsohn-Zamir, D. 2012. Taking outcomes seriously. *Utah Law Review* 2012, 861–902.
- Lewinsohn-Zamir, D. 2013. Can't buy me love: Monetary versus in-kind remedies. *Illinois Law Review*, no. 1, 151.
- Meyerson, M. 1990. The efficient consumer form contract: Law and economics meets the real world. *Georgia Law Review* 24, 583–628.
- Miller v. Byrne, 916 P.2d 566 (Colo.App.1995). New Jersey court overrules small print in policy, *Jury Verdicts Weekly*. (Supp. to Jury Verdicts Weekly). 3. (1969).
- Nisbett, R., and L. Ross. 1980. *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.
- Perillo, J. H. 2000. Misreading Oliver Wendell Holmes on Efficient Breach and Tortious Interference, 68 *Fordham L. Rev* 1068, at 1068.
- Pollock, F., and F. W. Maitland. 1898. *The history of English law before the time of Edward I*. 2nd ed. Cambridge: University Press.
- Posner, R. 1972. Economic analysis of law. *Restatement, Second, of Contract*. 1st ed. Boston: Little, Brown and Co.



- Skilton, R., and O. Helstad. 1967. Protection of the installment buyer of goods under the uniform commercial code. *Michigan Law Review* 65, 1465, 1476.
- Sunder, S. 2012. When and why individuals obey form-adhesive contracts: Experimental evidence of consent, compliance, promise, and performance. *Journal of Institutional and Theoretical Economics* 168(1), 147–49.
- Teichman, D. 2010. Old habits are hard to change: A case study of Israeli real estate contracts. *Law & Society Review* 44(2), 299–330.
- Thayer, J. B. 1898. *A preliminary treatise on evidence at the common law*. Boston: Little, Brown, and Co.
- Trebilcock, M. J. 1993. *The limits of freedom of contract*. Cambridge: Harvard University Press.
- Tversky, A., and Kahneman, D. 1981. The framing of decisions and the psychology of choice. *Science* ns. 211(4481), 453–58.
- Tversky, A., and Kahneman, D. 1982. Availability: A heuristic for judging frequency and probability. D. Kahneman, P. Slovic, and A. Tversky, eds., *Judgement under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- Tversky, A., and Kahneman, D. 1986. Rational choice and the framing of decisions. *Journal of Business* 59(S4), S251, S252.
- Vanberg, C. 2008. Why do people keep their promises? An experimental test of two explanations. *Econometrica* 76(6), 1467–80.
- Viscusi, W. and W. Magat. 1987. *Learning about risk: Consumer and worker responses to hazard information*. Cambridge: Harvard University Press.
- Weinstein, N. 1980. Unrealistic optimism about future life events. *Journal of Personality and Social Psychology* 39, 806–20.
- Wilkinson-Ryan, T. 2010. Do liquidated damages encourage breach? A psychological experiment. *Michigan Law Review* 108(5), 633–72.
- Wilkinson-Ryan, T. 2011. Breaching the mortgage contract: the behavioral economics of strategic default. *Vanderbilt Law Review* 64(5), 1547–83.
- Wilkinson-Ryan, T. 2012. Legal promise and psychological contract. *Wake Forest Law Review* 47, 843–74.
- Wilkinson-Ryan, T., and J. Baron. 2009. Moral judgment and moral heuristics in breach of contract. *Journal of Empirical Legal Studies* 6, 405–23.
- Wilkinson-Ryan, T., and D. Hoffman. 2010. Breach is for suckers. *Vanderbilt Law Review* 63, 1003–45.
- Williamson, O. E. 1975. *Markets and hierarchies, analysis and antitrust implications: A study in the economics of internal organization*. New York: Free Press.
- Wladis, J. D. 1987. Common law and uncommon events: The development of the doctrine of impossibility of performance in English contract law. *Georgetown Law Journal* 75, 1575–1631.
- Zamir, E. 2012. Loss aversion and the law. *Vanderbilt Law Review* 65(3), 829–94.

## CHAPTER 18

---

# CONSUMER TRANSACTIONS

---

OREN BAR-GILL

### 1 INTRODUCTION

---

CONSUMER transactions are ubiquitous. Consumers routinely enter into contracts with providers of goods and services—from credit cards and mortgages to cellphones, cable TV and Internet services to household appliances and theater and sports events to health club and magazine subscriptions, and more.

Consumer transactions differ from the archetypal arm's-length contract on which both classical contract law and neoclassical economics focus. Indeed, a defining feature of consumer transactions is the imbalance—with respect to both information and sophistication—between sellers and buyers. Consumer contracts are not the product of negotiations between two parties. Rather, these contracts are unilaterally drafted by sellers and mass-marketed to consumers, together with the products and services that these sellers offer.

A major concern about consumer contracts is that they include a host of fine-print terms, that consumers do not read these fine-print terms, and that sellers hide unfair and inefficient terms in the maze of fine print (Hillman and Rachlinski 2002; Korobkin 2003; Bakos, Marotta-Wurgler and Trossen 2009; Radin 2012). But concern about consumer transactions is not limited to the fine print. Even pricing schemes and other seemingly salient dimensions of the consumer transaction are often designed in a way that hurts consumers and reduces social welfare. When designing these non-fine-print terms, sellers cannot ignore the preferences and perceptions of consumers, even when the contract is offered to the consumer on a take-it-or-leave-it basis. If a seller fails to offer an attractive product-contract bundle, the consumer would purchase from another seller who does.

The underlying concern is that, even beyond the fine print, imperfectly informed and imperfectly rational consumers might fail to fully comprehend the costs and benefits of the product or service that they are purchasing. Sophisticated sellers can be expected to design their products, prices, and contracts in response to consumer misperception. Indeed, the design of consumer transactions can best be viewed as the outcome of an interaction between market forces and the psychology of consumers.

The remainder of this chapter is organized as follows. Section 2 conducts a positive, or descriptive, analysis of transacting in consumer markets. It shows how the interaction between market forces and consumer psychology results in a behavioral market failure. Section 3 explores the welfare implications of the behavioral market failure. Section 4 discusses possible legal policy responses. Section 5 offers a brief conclusion.<sup>1</sup>

## 2 A BEHAVIORAL-ECONOMICS THEORY OF CONSUMER TRANSACTIONS

---

There are two sides to the typical consumer transaction: On the demand side, there is an imperfectly rational consumer whose purchasing decisions are affected by systematic misperceptions.<sup>2</sup> On the supply side, there is a sophisticated seller who designs its products, contracts, and prices in response to consumers' misperceptions.<sup>3</sup> The consumer transaction, and specifically the design of contracts and prices, can be understood as the product of an interaction between consumer psychology and market forces.

Market forces demand that sellers be attentive to consumer psychology. Sellers who ignore consumer biases and misperceptions will lose business and forfeit revenue and profits. Over time, the sellers who remain in the market, profitably, will be the ones who have adapted their contracts and prices to respond, in the most optimal way, to

<sup>1</sup> Important contributions in the burgeoning field of behavioral industrial organization consider questions of contract design, albeit with less emphasis on policy implications. For an excellent recent textbook that summarizes and synthesizes this important literature, see Spiegler (2011). Spiegler's text, and the economic theory literature that it summarizes, is of foundational importance to many of the themes explored in this chapter. Throughout the chapter, I often cite Spiegler, rather than the original research papers on which he relies. For an earlier survey of the economics literature, see Ellison (2006).

<sup>2</sup> That individual decisionmaking is affected by a myriad of biases and misperceptions is well documented, see, e.g., Kahneman, Slovic, and Tversky (1982).

<sup>3</sup> Obviously, not all sellers are perfectly rational. Still, many sellers are sufficiently sophisticated to identify systematic consumer misperceptions and to design their products, contracts, and prices in response to these misperceptions. Alternatively, sellers that *happen to* offer products, contracts, and prices that optimally respond to consumer misperception will thrive in a competitive market at the expense of other sellers. As a result, sellers that survive and prosper in the marketplace will behave "as if" they are sufficiently sophisticated to identify systematic consumer misperceptions and to design their products, contracts, and prices in response to these misperceptions.

the psychology of their customers. This general argument is developed in section 2.1 below. In particular, the interaction between consumer psychology and market forces results in two common contract design features: complexity and cost deferral. Section 2.2 describes these features and explains why they appear in many consumer contracts. (For an excellent exposition of the economic theory literature on the topics addressed in section 2, see Spiegler 2011; Armstrong 2008.)

## 2.1 Maximizing the Perceived (Net) Benefit from a Transaction

### 2.1.1 A Behavioral Market Failure

When consumers are perfectly informed and perfectly rational, sellers design their products, contracts, and pricing schemes to maximize the net benefit to consumers. Otherwise they will lose business.<sup>4</sup> Specifically, sellers seek to maximize the actual (net) benefits to the consumer, which equals the actual benefit that the consumer gains from the product minus the actual price that the consumer pays for the product.

When consumers are imperfectly informed and imperfectly rational and, as a result, misperceive benefits and prices, sellers design their products, contracts, and pricing schemes to maximize the *perceived* (net) benefit to consumers, which equals the benefit that the consumer thinks she will gain from the product minus the price that the consumer thinks she will pay for the product.

It is not hard to see why products, contracts, and prices that are designed to maximize the perceived (net) benefit to consumers will generally look very different from products, contracts, and prices that are designed to maximize the actual (net) benefit. Consider the following example: A consumer is searching for a credit card. The consumer will borrow \$1,000 on her credit card in the first year and \$1,000 in the second year. (Assume, for simplicity, that the size of the credit card debt is independent of the interest rate charged, and that there is no time discounting.) The (risk-adjusted) cost of funds to the issuer, that is, the cost of providing the credit card loan, is 10% annually. For the particular consumer, this translates into a total cost of \$200 ( $= 10\% \times \$1,000 + 10\% \times \$1,000$ ).

If the consumer is rational and accurately anticipates the extent of borrowing, then a credit card issuer, operating in a competitive market, would set an interest rate of 10% for the first year and the same interest rate, 10%, for the second year. Denote this pricing scheme as the (10%, 10%) contract. The (10%, 10%) contract reflects per-period, marginal-cost pricing, which will be observed in a competitive market. With this contract, the consumer will pay interest of \$100 ( $= 10\% \times \$1,000$ ) in the first year and \$100 ( $= 10\% \times \$1,000$ ) in the second year, for a total amount of \$200; and the issuer will just cover her costs.

<sup>4</sup> A monopolist will design efficient products and contracts that maximize the (gross) benefit enjoyed by the informed, rational consumer. The monopolist, however, will set higher prices to extract much of this benefit.

But what if the consumer is imperfectly rational? In particular, what if this imperfectly rational consumer, while realizing that he will borrow \$1,000 in the first year, falsely believes (perhaps as a result of overoptimism—see section 2.2.2 below) that he will not need to resort to credit card debt in the second year? A sophisticated issuer, who understands that the consumer underestimates his second-year borrowing, will no longer offer the (10%, 10%) contract. For the imperfectly rational consumer, the (10%, 10%) contract implies a perceived total interest payment of \$100 ( $= 10\% \times \$1,000 + 10\% \times \$0$ ). By altering the contractual design, the issuer can further reduce the perceived total interest.

Consider a (5%, 15%) contract, setting a low, 5% interest rate for the first year and a high, 15% interest rate for the second year. The issuer still covers her costs of \$200 ( $= 5\% \times \$1,000 + 15\% \times \$1,000$ ). And the consumer perceives a lower total interest payment of \$50 ( $= 5\% \times \$1,000 + 15\% \times \$0$ ). The altered contract design reduces the perceived total price (or, equivalently, increases the perceived net benefit) of the loan, in the eyes of the imperfectly rational consumer. Sophisticated issuers will understand this and offer the (5%, 15%) contract, instead of the (10%, 10%) contract. Indeed, teaser rate contracts, with low interest rates for an initial, introductory period and higher long-term rates, are common in the credit card market.

Before we proceed further, the relationship between imperfect information and imperfect rationality should be clarified. Rational-choice theory allows for imperfect information. A divergence between perceived benefits and prices on the one hand and actual benefits and prices on the other is also possible in a rational-choice framework with imperfectly informed consumers. The focus here, however, is on systemic under- and overestimation of benefits and prices. Perfectly rational consumers will not have systemically biased beliefs; imperfectly rational consumers will. The main difference is in how perfectly and imperfectly rational consumers deal with imperfect information. Rational-choice decision-making provides tools for effectively coping with imperfect information. These tools are not used by the imperfectly rational consumer. Instead, he relies on heuristics or cognitive rules-of-thumb, which result in predictable, systemic biases and misperceptions. Moreover, while the perfectly rational consumer realizes that she is imperfectly informed, the imperfectly rational consumer might be blissfully unaware of the extent of his ignorance.

### 2.1.2 *The Role of Competition*

Sophisticated sellers design the consumer transaction in response to the cognitive biases and misperceptions of their customers. As we will see, when contract design responds to consumer psychology rather than sellers' cost structure, the resulting distortions reduce welfare and hurt consumers. Can enhanced competition ameliorate, or mitigate, this behavioral market failure?

The first-cut answer is no. Continuing with the credit card example, consider two competing issuers—a high-road issuer and a low-road issuer. The high-road issuer does not exploit the cognitive biases of its customers. It thus offers the (10%, 10%) contract. The low-road issuer, on the other hand, has no qualms about exploiting the

cognitive biases of its customers. It thus offers the (5%, 15%) contract, which, in the eyes of the biased consumer, appears to be more attractive. If many consumers are imperfectly rational, the high-road seller will lose out to the low-road seller. These consumers will flock to the low-road seller and the high-road seller will go out of business.

Competition does not solve the behavioral market failure. The reason is straightforward: Competition forces sellers to maximize the perceived (net) consumer benefit. When consumers accurately perceive their benefits, competition will help consumers. But when consumers are imperfectly rational, competition will maximize the perceived (net) benefit at the expense of the actual (net) benefit. Focusing on price: When consumers are perfectly rational, sellers compete by offering a lower price. When consumers are imperfectly rational, sellers compete by designing pricing schemes that create an appearance of a lower price. The underlying problem is on the demand side of the market: imperfectly rational consumers generate biased demand. Competition forces sellers to cater to this biased demand.

Modern, neoclassical economics recognizes that even perfectly competitive markets can fail, because of externalities and asymmetric information. Behavioral economics adds a third cause for market failure: misperception and bias. This behavioral market failure is a direct extension of the imperfect information problem. Rational consumers form unbiased estimates of imperfectly known values. Faced with similarly limited information, imperfectly rational consumers form biased estimates. Unbiased estimates can cause market failure; biased estimates can cause market failure.

### 2.1.3 *Market Correction*

The preceding analysis takes consumers' biases and misperceptions as exogenously given. With exogenous biases and misperceptions, competition does not ameliorate the behavioral market failure. Indeed, competition forces sellers to exploit the cognitive biases of their customers. But perceptions and misperceptions can be endogenous. In particular, sellers can influence consumer perceptions, for example, through marketing. With endogenous perceptions, sellers, operating in a competitive market, might try to exacerbate biases that increase the perceived benefit and reduce the perceived price of their products (see, e.g., Glaeser 2004). But sellers may also offer superior, yet underappreciated products and contracts and try to compete by educating consumers and fighting misperception (see, e.g., Gillette 2004).

It is not clear a priori whether sellers will compete by exacerbating consumer biases (or simply taking them as given) or by trying to fight these biases. There are, however, several forces working against the more optimistic bias-correction alternative: First, bias-correction suffers from a collective-action problem. If seller A succeeds in educating consumers about the risks of a certain contract design, and the benefits of its alternative design, then other sellers will copy seller A's design and compete away any profit that seller A could have made. Anticipating such copying, and loss of profit, seller A would be reluctant to invest the substantial resources needed to educate consumers (Beales, Craswell, and Salop 1981).

In some markets, the collective-action problem is avoided by a first-mover advantage, which would enable seller A to earn sufficient profits, before other sellers can copy, to make the initial investment in consumer education worthwhile. Unfortunately, seller A is unlikely to enjoy a large first-mover advantage with contract design innovations. To replicate an improvement in a physical product, competitors need to reconfigure assembly lines. This takes time. To replicate a contract-design innovation, however, competitors only need to type and print, or upload on a website, a new contract.

Copying by competing sellers can prevent bias-correction, as it reduces the profit that seller A can expect to make from consumer education. In other cases, the profit from consumer education is small, even in the absence of copying. Consider cases where consumer misperception is not about a product attribute (e.g., a certain contract design), but rather about product use, that is, about how the consumer will use the product. Recall, in the credit card example, the consumer thought that she would use the credit card to borrow only in the first year, while in fact she used it to borrow in both the first and second years.

Seller A has a stronger incentive to correct product attribute mistakes, since the educated consumers will be attracted to seller A's superior product, at least until other sellers copy. Seller A has a weaker incentive to correct product use mistakes. A consumer who learns that she will in fact borrow in both periods might as well get a credit card from a competing seller. Product attribute information is seller-specific, and so the seller who discloses this information, or educates consumers about this information, will enjoy a competitive advantage. Product use information, on the other hand, is consumer-specific, and so the disclosing seller enjoys no competitive advantage (Bar-Gill and Board 2012).

Yet another force reduces sellers' incentives to educate consumers in markets with add-on products. Think of a hotel that, in addition to the basic room rate, charges for add-ons like wireless Internet, room service, and pay-per-view TV. If imperfectly rational consumers, when searching for a hotel, do not focus on the add-ons, sellers will set a lower price for the room and high prices for the add-ons. Now consider a high-road hotel that invests in educating consumers about the importance of add-ons. Some of the newly educated consumers may choose the high-road hotel with the higher room rates and lower add-on prices. But others will still prefer the low-road competitor with the low room rates; they will use their newfound knowledge about add-ons to plan ahead and avoid the high add-on prices—by eating at a local restaurant, rather than ordering room service, or by watching a movie on their tablet rather than on the hotel TV. As a result, the high-road hotel might not find it worthwhile to educate consumers about add-on prices. (See Gabaix and Laibson 2006.)

To sum up, sellers, operating in a competitive market, may choose a bias-correction strategy in certain cases. In many other cases, however, they will choose a bias-exploiting strategy. We cannot always count on competition to ameliorate the behavioral market failure.

Before moving on, another market correction force—or set of market correction forces—should be mentioned. Consumers, even imperfectly rational consumers, learn



from their mistakes, and from the mistakes of others. Such learning could, in principle, solve the behavioral market failure or, at least, mitigate it. Expert advice and seller reputation can further facilitate mistake correction. While clearly valuable, these market solutions are imperfect. Learning occurs more quickly in some contexts, but less quickly in others. And when the harm is substantial, for example, when a consumer's home is foreclosed because she failed to fully understand the terms of her mortgage, learning may come too late. Expert advice can be helpful, but such advice is not always sought out and it is not always reliable. And reputation depends on the effective and accurate flow of information among consumers, which sometimes occurs, but not always. In many consumer markets, consumer misperception will persist and sellers will continue to design their products, contracts, and prices in response to these misperceptions. (See Bar-Gill 2012 and references cited therein.)

## 2.2 Common Contract Design Features

Sellers design consumer transactions—products, contracts, and prices—in response to the imperfect rationality of their customers. This behavioral market failure manifests in two common contract design features: complexity and deferred costs. (Complexity and deferred costs are also recurring themes in theoretical models of industrial organization with imperfectly rational consumers—see Spiegler 2011, ch. 12; see also Armstrong 2008.)

### 2.2.1 *Complexity*

Consumer contracts are complex! Just look at your credit card contract, cell phone contract, mortgage contract, checking account contract, insurance contract, and so on. There is a heap of fine print, full of technical legal language, contributing to this complexity. Moreover, substantial complexity is observed on the non-fine-print terms, most prominently the contract price. The pricing, in many consumer transactions, is multidimensional and complex. Credit cards come with many different interest rates and fees, calculated using complex formulas. The same is true for mortgage contracts, especially if one considers the subprime mortgages that contributed to the economic crises of 2008. Cell phone pricing is also notoriously complex, and there are many other examples.

There are efficiency justifications for multidimensional pricing. Compare a simple credit card contract with only an annual fee and a basic interest rate for purchases to a complex credit card contract that, on top of these two price dimensions, adds a default interest rate, a late fee, and a cash advance fee. The complex card facilitates risk-based pricing and tailoring of optional services to heterogeneous consumer needs. The default interest rate and the late fee allow the issuer to increase the price for consumers who, after the initial contract is entered into, reveal themselves to be higher-risk borrowers, for example, by paying late. Such efficient risk-based pricing is impossible with the simple card. The single interest rate design does not allow for price adjustments

in response to new information about borrower risk, resulting in cross-subsidization of high-risk consumers by low-risk consumers. Similarly, the cash advance fee allows the issuer to charge separately for cash-advance services, which benefit some consumers but not others. Such tailoring of optional services is impossible with the simple card. An issuer using the simple contract design will set a higher annual fee, to cover the cost of providing cash-advance services, and, as a result, consumers who use the cash-advance feature will be cross-subsidized by those who do not. (See, e.g., Bar-Gill and Bubb 2012.)

These efficiency benefits explain some of the complexity and multidimensionality observed in consumer contracts. But they cannot explain all of the staggering complexity that consumers face. There is another, behavioral explanation. Complexity hides the true cost of the product from the imperfectly rational consumer. A rational consumer navigates complexity with ease. She assesses the probability of triggering each rate, fee, and penalty and calculates the expected cost associated with each price dimension. The rational consumer may have imperfect information, but she will form unbiased estimates given the information that she chose to collect. Accordingly, each price dimension will be afforded the appropriate weight in the overall evaluation of the product.

The imperfectly rational consumer, on the other hand, is incapable of such an accurate assessment. He is unable to calculate prices that are not directly specified. Even if he could perform this calculation, he would be unable to simultaneously consider multiple price dimensions. And even if he could recall all the price dimensions, he would be unable to calculate the impact of these prices on the total cost of the product. The imperfectly rational borrower deals with complexity by ignoring it. He simplifies his decision problem by overlooking nonsalient price dimensions (see Thaler 1999). And he approximates, rather than calculates, the impact of the salient dimensions that cannot be ignored. In particular, limited attention and limited memory result in the exclusion of certain price dimensions from consideration. Limited processing ability prevents borrowers from accurately aggregating the different price components into a single, total expected price that would serve as the basis for choosing the optimal product. While the rational consumer is unfazed by complexity, the imperfectly rational consumer might be misled by complexity. (See Bar-Gill 2012; Korobkin 2003.)

As explained above, when consumers are imperfectly rational, sellers design contracts in response to systematic biases and misperceptions. In particular, they reduce the total price as perceived by consumers by increasing nonsalient prices and decreasing salient prices. This strategy depends on the existence of nonsalient prices. In a simple contract, the one or two price dimensions will generally be salient. Only a complex contract will have both salient and nonsalient price dimensions. Complexity thus serves as a tool for reducing the perceived total price.

Back to the nonprice, legal terms: Sellers often include a long list of fine-print terms in their consumer contracts. Common examples include arbitration clauses, forum selection clauses, choice of law clauses, and liability waivers. These all add to the complexity of the consumer contract. Consumers do not read the fine print (Bakos, Marotta-Wurgler, and Trossen 2009). And since these terms are unlikely to

be salient to many consumers, they will often be proseller. (But see Marotta-Wurgler 2007a, 2007b.) One type of legal term deserves special mention: the unilateral change clause. This term allows, or at least purports to allow, the seller to modify the contract unilaterally at any time and for any reason. (Sellers may be able to modify their contracts, in an effectively unilateral way, even without an explicit unilateral change clause.) The ability to change terms further increases the complexity of the consumer transaction. The concern, of course, is that these changes will often be proseller and that imperfectly rational consumers, when entering the transaction, will underestimate the potentially adverse implications of future changes in terms (Bar-Gill and Davis 2010).

A clarification is in order. In theory, an incomplete understanding of complex contracts is consistent with rational-choice theory. Facing a complex contract, a rational consumer would have to spend time reading the contract and deciphering its meaning. If the cost of attaining perfect information and perfect understanding of the contract is high, the rational borrower would stop short of this theoretical ideal. Imperfect rationality can be viewed as yet another cost of attaining more information and better understanding. When this cost component is added, the total cost of becoming informed goes up, and thus the consumer will end up with less information and a less complete understanding of the contract. Imperfect rationality, however, is not simply another cost component. A rational consumer who decides not to invest in reading and deciphering certain contractual provisions will not assume that these provisions are favorable to her. In fact, she will recognize that unread provisions will generally be proseller. In contrast, an imperfectly rational consumer will completely ignore the unread or forgotten terms or naively assume that they are favorable to him. Accordingly, a complex, unread term or a hidden fee would lead an imperfectly rational consumer—but not a rational consumer—to underestimate the total cost of the product. As a result, the incentive to increase complexity and hide fees will be stronger in a market with imperfectly rational consumers. The behavioral economics theory of contract design is an imperfect-rationality theory, not an imperfect-information theory.

### 2.2.2 *Deferred Costs*

Nonsalient price dimensions and prices that impose underestimated costs create opportunities for sellers to reduce the perceived total price of their product. What makes a price nonsalient? What leads consumers to underestimate the cost associated with a certain price dimension? While there is no simple answer to these questions, there is one factor that exerts substantial influence on salience and perception—time.

The basic claim is that, in many cases, noncontingent, short-run costs are accurately perceived, while contingent, long-run costs are underestimated. Take the credit card example: an annual fee is to be paid for certain and soon. This cost will figure prominently, when the consumer chooses among competing cards. A late fee is to be paid in the future and only if the consumer makes a late payment. This cost will often be

underestimated by the consumer. It is less likely to affect card choice.<sup>5</sup> If costs in the present are accurately perceived and future costs are underestimated, market forces will produce deferred-cost contracts.

Many fine-print, legal terms are also examples of cost deferral. These terms often address future breach contingencies. Liability waivers fall into this category, as do dispute resolution terms, like arbitration clauses and choice of law provisions. Consumers who underestimate the probability of breach will underestimate the importance of these terms. The unilateral change clause is similarly forward looking. Consumers who underestimate the incidence of contingencies that might trigger a change of terms will also underestimate the cost that a unilateral change clause imposes on them.

The importance of the temporal dimension of price and cost can often be traced back to two underlying forces: myopia and overoptimism. Myopic consumers care more about the present and not enough about the future. People are impatient—they prefer immediate benefits, even at the expense of future costs (see, e.g., Loewenstein and O'Donoghue 2004; O'Donoghue and Rabin 1999). Myopia is attributed to the triumph of the affective system, which is driven primarily by short-term payoffs, over the deliberative system, which cares about both short-term and longer-term payoffs. This understanding of myopia, and of intertemporal choice more generally, is consistent with findings from neuroscience (see Loewenstein and O'Donoghue 2004; McClure et al. 2004).

In addition, future costs are often underestimated, because consumers are overoptimistic. The prevalence of the optimism bias has been confirmed in multiple studies (see, e.g., Weinstein 1980; Svenson 1981; see also chapter 13 by Williams in this volume). Optimistic consumers tend to underestimate the probability of triggering contingent, future costs. They underestimate the likelihood that the contingency will materialize. For example, an optimistic cardholder might underestimate the probability of making a late payment, leading her to underestimate the importance of the late fee. Similarly, when mortgage contracts set low introductory interest rates coupled with high long-term rates, overoptimism may cause the consumer to underestimate the importance of the high long-term interest rates. The optimist might overestimate the probability of exiting the mortgage contract before the high rates kick in, by selling the house or refinancing the mortgage, because she underestimates the probability that falling real estate prices will make it difficult to sell or refinance.

A sophisticated seller facing imperfectly rational consumers will seek to reduce the perceived total price of her product without reducing the actual total price that consumers pay. When consumers are myopic or optimistic, this wedge between perceived and actual prices can be achieved by backloading costs onto long-term price dimensions. The result is deferred-cost contracts.

<sup>5</sup> Even late fees are becoming salient, as evidenced by recent ads that emphasize the no-late-fee feature of certain card products. See, e.g., Citi's Simplicity Card and Discover's It Card. For a more comprehensive analysis of the dynamics of salience in the credit card market, see Bar-Gill (2012, ch. 2).

## 3 WELFARE IMPLICATIONS

---

Consumer markets often suffer from a behavioral market failure. What are the welfare implications of this market failure? Specifically, what are the consequences of complexity and deferred costs—when responding to imperfect rationality—for consumers and for efficiency?

### 3.1 Hindered Competition

The excessive complexity of many consumer contracts hurts competition to the detriment of consumers and the economy at large. For competition to work well, consumers must be able to compare the benefits and costs of different products and choose the one that provides the best value, given the consumer's tastes and needs. Gathering information on competing products is costly, and complexity—of the product or contract—increases this cost. When the cost of collecting information goes up, the consumer will collect less information. This is true for the rational consumer. It is even more true for the imperfectly rational consumer, who might be effectively paralyzed by the complexity.

Less information, and less comparison shopping, imply weaker competition. Sellers gain market power, increasing profits at the expense of consumers. Limited competition also imposes a welfare cost in the form of inefficient allocation, as consumers are not matched with the most efficient seller. (A series of recent papers in industrial organization argue that firms introduce spurious complexity into tariff structures and by doing so inhibit competition and reduce welfare. See, e.g., Ellison 2005; Gabaix and Laibson 2006; Spiegler 2006; Ellison and Ellison 2009. For a discussion of additional welfare implications of complex contracts, see Gilo and Porat 2006.)

### 3.2 Distorted Competition

Complexity weakens the forces of competition. But even if sellers vigorously competed for consumers, biases and misperceptions on the demand side of the market would distort these competitive efforts, leading to suboptimal outcomes for consumers and reducing social welfare. As explained above, sellers try to maximize the perceived net benefit of their products in the eyes of consumers. When consumer perceptions are biased, the products, contracts, and prices that maximize the perceived net benefit are different from those that maximize the actual net benefit. The result is a distorted contract design, with excessive complexity and deferred costs.

Focusing on price, sellers facing rational consumers will try to minimize the total price of their product. Competition would operate on the total-price level. Imperfectly

rational consumers, on the other hand, choose products based on a few salient price dimensions. Competition will thus focus on those salient prices, driving them down, while nonsalient prices, free from competitive pressure, increase. And when salience is a function of time—when short-term prices are salient and long-term prices are not—competition will drive short-term prices below cost, with sellers recouping losses through high long-term prices.

Such pricing entails two types of efficiency costs. First, product-use decisions will be distorted. Prices affect product-use decisions. A high late fee deters late payments. A low introductory interest rate induces borrowing during the introductory period. Optimal pricing provides accurate incentives: With an optimal late fee, consumers will pay late if and only if the benefit of paying late outweighs the cost of late payment (including the added risk implied by late payment) to the issuer. With an optimal interest rate, consumers will borrow if and only if the benefit from borrowing outweighs the issuer's cost of providing credit. Optimal price tracks the seller's cost so that consumers pay the price and use the product only when the benefit to them outweighs the seller's cost.

Second, salience-based pricing distorts product choice: Sellers reduce salient prices and increase nonsalient prices in order to minimize the total price as perceived by the imperfectly rational consumer. Since the perceived total price will be lower than the actual total price, biased consumers may well choose a product that costs more than it is worth to them. The result is inefficient allocation.

This inefficiency exists even with optimal pricing. Here, the nonsalient price dimensions will be ignored or underestimated, reducing the perceived total price. Distorted contract design exacerbates the problem by backloading more of the total price onto the nonsalient, underestimated dimensions. The gap between actual total price and perceived total price increases, as does the number of consumers who purchase products that reduce their welfare. Bias and misperception result in artificially inflated demand. Distorted contract design adds air to the demand balloon.

### 3.3 Distributional Concerns

The distributional implications of complexity and deferred costs are, in large part, market specific. Still, a few general observations can be made: Excessive complexity imposes a larger burden on less sophisticated consumers and on financially weaker consumers who cannot hire advisers to help them navigate the complexity of products and contracts. Thus, complexity has a regressive distributional effect.

As for deferred costs: Distorted pricing—specifically, low salient or short-term prices and high nonsalient or long-term prices—shifts the burden to the group of consumers who are more likely to pay the high, nonsalient prices. In some cases, such as when the nonsalient prices are default or penalty prices (late fees and default interest rates,

for instance), weaker consumers are more likely to shoulder the burden of the high, nonsalient prices. In these cases, the deferred-cost feature will have a regressive distributional effect.

## 4 POLICY IMPLICATIONS

---

The identification of a market failure—the behavioral market failure—that results in possibly substantial welfare costs opens the door for considering the potential role of legal policy. What can the law do to help consumers and enhance efficiency? Several main legal tools or regulatory techniques are discussed below—starting with hard paternalistic policies, in section 4.1, and then focusing on soft paternalistic policies in the remaining sections—disclosure regulation in section 4.2, default rules and safe harbors in section 4.3, and the right to withdraw from the transaction in section 4.4. (On the range of policy choices, and their normative evaluation, see chapter 28 by Sunstein in this volume; Sunstein, 2013; Bar-Gill and Ben-Shahar 2013.)

### 4.1 Mandatory Rules

When a feature of a consumer transaction—a feature designed in response to bias or misperception—is found to hurt consumers and reduce social welfare, a natural response would be to ban the feature, that is, to prohibit sellers from using the specific contract term or practice. There are many examples of such prohibitions—limits on late fees in the credit card market, restrictions on prepayment penalties in the mortgage market, restrictions on early termination fees in the cell phone market, and more.<sup>6</sup>

In some cases legislators and regulators impose specific bans, targeting particular contract terms and practices in specific markets. In other cases, courts and agencies use vague standards—from general legislation or from the common law—to strike down different terms and practices across different markets.<sup>7</sup>

<sup>6</sup> The Credit Card Accountability, Responsibility and Disclosure Act of 2009 (CARD Act) bans certain payment allocation methods used by card issuers and restricts the magnitudes of certain fees. The use of prepayment penalties has been substantially curtailed by the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010. Some courts have stricken early termination fees in cell phone contracts. See *In re Cellphone Termination Fee Cases*, 193 Cal App 4th 298, 122 Cal Rptr 3d 726 (2011), reh'g denied (Mar. 24, 2011), review denied (June 15, 2011), *cert denied*, 132 S Ct 555, 181 L Ed 2d 397 (2011).

<sup>7</sup> The restrictions imposed by the CARD Act and by the Dodd-Frank Act are examples of specific bans targeting particular terms and practices. The Cellphone Termination Fee cases provide an example where a vague standard was used to police consumer contracts. Other examples of courts using vague standards include the use of the unconscionability doctrine to strike down certain



These bans and prohibitions are examples of hard paternalism. While often effective in eliminating the targeted term or practice, hard paternalism comes at a price. Consumers are heterogeneous, with different needs and preferences and with different degrees of sophistication. A practice that harms one, less sophisticated consumer may benefit another, more sophisticated consumer. For instance, high late fees might harm less sophisticated consumers who underestimate the impact of these fees. At the same time, these fees can enhance efficiency—by facilitating risk-based pricing—and thus benefit other, more sophisticated consumers. (See, e.g., Bar-Gill and Bubb 2012.)

This cost of hard paternalism raises two questions. First, can we be certain that the cost of the legal intervention does not exceed the benefit from the intervention? In many cases, lawmakers will not have sufficient information to conduct the necessary cost-benefit analysis.<sup>8</sup> Second, even if the benefit clearly exceeds the cost, is it necessary to bear the cost? In other words, is there an alternative mode of legal intervention that can secure the same (or similar) benefit at a lower cost? The next sections explore such alternative modes of legal intervention, which collectively fall under the heading of “soft paternalism.” These regulatory techniques aim to minimize any interference with consumer autonomy and market forces. They strive to help the less sophisticated consumer, while imposing minimal costs on the more sophisticated consumer. (See Thaler and Sunstein 2008; Sunstein and Thaler 2003; Thaler and Sunstein 2003; Camerer et al. 2003.)

## 4.2 Disclosure Regulation

One of the central tools in the soft paternalism arsenal is disclosure regulation.<sup>9</sup> By requiring sellers to disclose information, the lawmaker can enhance consumer autonomy and facilitate the operation of market forces. The uninformed, less sophisticated consumer benefits, while the more sophisticated consumer is not harmed.<sup>10</sup> Disclosure

one-sided arbitration clauses in credit card contracts. (See *Discover Bank v. Superior Court*, 30 Cal. Rptr.3d 76 (2005); but see *AT&T Mobility v. Concepcion*, 563 U.S. 321 (2011), overruling *Discover Bank*.) The Federal Trade Commission has invoked its authority under Section 5 of the Federal Trade Commission Act to ban a series of practices that it deemed unfair and/or deceptive. And courts have used state-level Unfair and Deceptive Acts and Practices statutes to attack a broad range of terms and practices. Russell Korobkin argued that unconscionability doctrine—another vague standard—should be used to address the market failure caused by consumers’ imperfect rationality (Korobkin 2003).

<sup>8</sup> An imperfectly informed lawmaker could still produce estimates of the costs and benefits of the legal intervention and decide to intervene if the benefit estimate exceeds the cost estimate. The lawmaker will recognize, however, that there is a positive probability that her estimates are inaccurate and that the true cost exceeds the true benefit, such that the legal intervention results in a net welfare loss, from an ex post perspective.

<sup>9</sup> Some commentators view disclosure mandates as not paternalistic at all (rather than “softly” paternalistic). Yet disclosure mandates can be more paternalistic than they appear. See Bar-Gill (2012).

<sup>10</sup> There could be some indirect harm, as sellers pass on some of the cost of compliance with the disclosure mandates to consumers through higher prices.

mandates can be an effective response to imperfect information when consumers are perfectly rational. Disclosure regulation can be similarly important, and perhaps even more important, when consumers are imperfectly rational. Moreover, the optimal design of disclosure mandates must take into account the imperfect rationality of consumers, as detailed below.<sup>11</sup>

#### 4.2.1 *Two Categories of Information*

When designing disclosure mandates, lawmakers should consider what type of information sellers should disclose. It is useful to distinguish between product-attribute information and product-use information. Product-attribute information includes information on what the product is and what it does: product features, contract terms, and price terms. Product-use information includes information on how the consumer will use the product, for example, will the consumer talk more or less on her cell phone, how much will the consumer borrow on his credit card, and so on. Product use depends on product attributes (e.g., consumers will borrow less on their credit card when the interest rate is higher), but it also depends on the consumer's needs and preferences. As a general matter, both types of information are important; they are both necessary if the consumer is to make optimal decisions about product choice and product use.

#### 4.2.2 *Disclosing Product-Attribute Information*

Disclosure mandates have traditionally focused on product-attribute information (Bar-Gill and Ferrari 2010). Sellers have better product-attribute information and, thus, it makes sense to require that they disclose this information to consumers. Such disclosure would reduce the problem of information asymmetry and enhance market efficiency. In certain contexts, disclosure mandates have been shown to be effective. Specifically, evidence suggests that the APR disclosure has been successful, at least to some extent, in the credit card and mortgage markets (Bar-Gill 2012).

But in many other contexts mandatory disclosure of product-attribute information has not produced any tangible benefits. The disclosures are often too long and too complicated to be digested by the average, imperfectly rational consumer. Moreover, since disclosure mandates have been the favorite regulatory tool across multiple markets, the information overload problem far exceeds what could be attributed to any single disclosure form. (See Ben-Shahar and Schneider 2014 and the references cited therein; Marotta-Wurgler 2011.)

<sup>11</sup> The theory of optimal disclosure design is still not well developed. Most disclosure mandates are issued without any attempt to devise optimal disclosure forms in a scientific manner. In recent years, such regulators as the Federal Trade Commission, the Federal Reserve Board, and the Consumer Financial Protection Bureau have begun to employ consumer-testing methods to identify more effective disclosure forms. See Bar-Gill 2012, chs. 2 and 3. See also Sunstein (2010, p. 5) (emphasizing the importance of testing). Scholars have also begun to consider the question of optimal disclosure design. See, e.g., Craswell (2006).

### 4.2.3 *Disclosing Product-Use Information*

Disclosure mandates have, by and large, ignored product-use information. The implicit assumption is that consumers have better information about their own use patterns and thus it makes no sense to mandate disclosure of product-use information by sellers. This assumption, while valid in many consumer markets, fails in other important markets. In particular, this assumption fails in service markets, like the credit card and cell phone markets, where sellers maintain long-term relationships with their customers and collect large amounts of use information on each individual customer (see MacDonald 2007; Grubb 2009). In theory, consumers have all the use information that sellers have, but in practice consumers do not store, and do not remember, information on their past use. Sellers, on the other hand, store the information in large databases and analyze it using sophisticated algorithms (see Bar-Gill 2012). Accordingly, it may well make sense to require disclosure of product-use information. Indeed, evidence that consumers often do not have accurate perceptions about their use patterns, and that such misperceptions distort both product choice and product-use decisions, provides further impetus for mandating disclosure of product-use information (see Grubb 2009; Bar-Gill and Stone 2012).

There is another reason why lawmakers should consider mandating the disclosure of product-use information: sellers are less likely to voluntarily disclose such information. Sellers will often voluntarily disclose product-attribute information. Product-attribute information is, in many cases, seller-specific. A high-quality seller that discloses the superior attributes of her product will gain a competitive advantage. Standard unraveling results imply that much product attribute information will be voluntarily disclosed by sellers (see Grossman and Hart 1980; Grossman 1981; Milgrom 2008). The same is not true about product use information. While product-attribute information is seller-specific, product-use information is generally consumer-specific. Accordingly, a seller does not gain a competitive advantage by disclosing product-use information. If a cellphone company tells a customer exactly how much she uses her cell phone, the newly informed customer can purchase a plan that matches her use patterns from a competing seller. Since voluntary disclosure is less reliable in this context, it is more important to mandate the disclosure of product-use information (see Bar-Gill and Board 2012).

Product-use information can be disclosed at different levels. Ideally, individual use information should be disclosed. This is possible in service markets, like the credit card and cell phone markets, where sellers have long-term relationships with their customers and monitor the use-patterns of individual customers. In other markets, statistical use-pattern information can be disclosed. Specifically, sellers can disclose the use patterns of the average consumer (or the average consumer in a certain demographic subgroup). For instance, health clubs could disclose the average per-visit price of consumers who purchase an annual membership. Such a disclosure might lead some consumers to forgo the annual membership and choose a pay-as-you-go plan (see DellaVigna and Malmendier 2006).

Recent legislation and regulation have recognized the importance of product-use information. In the United States, the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 imposes a general duty, subject to rules prescribed by the new Consumer Financial Protection Bureau, to disclose information, including “usage data,” in markets for consumer financial products. Similarly, the Federal Communications Commission (FCC), in considering ways to improve disclosure in telecommunication service markets, noted also the disclosure of information on usage. In the U.K., the Department of Business Innovation and Skills and the Behavioral Insights Team in the Cabinet Office are promoting the MyData initiative, which aims to provide consumers with access to use information—their own use information—that sellers possess.<sup>12</sup> Academics have gone further, proposing mandatory disclosure of product-use information for credit cards, mortgages, payday loans, cell phones, subscription services, and more (see Nalebuff and Ayres 2003; Bar-Gill 2004; Lynch and Zauberman 2006; Sovern 2006; Thaler and Sunstein 2008; Bar-Gill and Stone 2009; Bar-Gill and Ferrari 2010; Kamenica, Mullainathan, and Thaler 2011).

While disclosing product-use information is important, it is also important to make sure that by disclosing this information, in addition to existing product-attribute information, we do not exacerbate the information overload problem (see Ben-Shahar and Schneider 2014). The challenge is to design disclosure that can improve the decision-making of imperfectly rational consumers. To meet this challenge disclosure regulation must take one of the following two approaches: (1) focus on simple disclosures for consumers, or (2) mandate comprehensive disclosures for intermediaries and sellers that will help consumers indirectly.

#### 4.2.4 *Simple Disclosures for Consumers*

If imperfectly rational consumers are to read and use disclosed information, the disclosure must be simple. The problem, of course, is that simple disclosures necessarily exclude some relevant information. The challenge is to provide as much information as possible in a disclosure that is simple enough to be useful to a majority of consumers.

In many cases, an effective way to provide the most information in the least complex way is by disclosing total-cost-of-ownership (TCO) information or annual cost information. The TCO (or annual cost) disclosure is a simple, single-figure disclosure that provides consumers with an estimate of how much they will ultimately pay for the product over the product’s life span (or over a one-year period).

The TCO and annual cost disclosures combine product-attribute information with product-use information. For example, the annual cost of a cell phone plan would combine plan rates with information on use patterns to estimate the annual cost of cellular service. For a new customer, the carrier may have to use average-use information to calculate the annual cost estimate. For existing customers, the carrier should use the individual-use information that it already has.

<sup>12</sup> See Pub. L. 111-203, Title X, Sec. 1033 (the Dodd-Frank Act); FCC (2009); Department of Business Innovation and Skills and Cabinet Office Behavioral Insights Team (2011).

These simple, aggregate disclosures, especially the TCO disclosure, are a direct response to the behavioral market failure identified in this chapter. Sellers facing imperfectly rational consumers will design complex, deferred-cost contracts in order to maximize the wedge between the actual and perceived cost of their products. The TCO disclosure undermines sellers' incentives to design such welfare-reducing contracts. Complexity is used to hide the true cost of the contract by allowing sellers to load costs onto less salient price dimensions. If sellers are required to provide a TCO disclosure that aggregates both salient and nonsalient prices, complexity ceases to be a problem for consumers and loses its appeal to sellers. Similarly, sellers design deferred-cost contracts so that myopic and overoptimistic consumers will underestimate the cost of the product. A TCO disclosure that aggregates both short-term and long-term costs into a single figure that guides consumer choice would substantially reduce sellers' incentives to defer costs.<sup>13</sup>

#### 4.2.5 *Comprehensive Disclosures for Intermediaries and Sellers*

The standard disclosure paradigm targets consumers; in other words, the disclosures are supposed to be read and used by consumers. But disclosures can also help consumers even when they are not targeted at consumers directly, but rather at intermediaries and even at other sellers.

Consider a consumer who is at the end of a two-year cellular service contract. This consumer needs to decide whether to stay with the current carrier and plan, or switch to a different plan with the same carrier, or switch to another carrier altogether. The consumer must choose among many complex products in the search for the optimal cell phone plan, given his or her particular use patterns. To do that, the consumer could employ the services of an intermediary. The intermediary will have information on available plans, that is, product-attribute information. But it will not have information on the consumer's use patterns. (Of course, the consumer could provide this information, but as suggested earlier, many consumers have a poor sense of their use patterns.) This missing information exists in the databases of the consumer's old carrier. Disclosure regulation could require carriers to provide this information, in electronic form, to the consumer. The consumer could then forward this data to the intermediary, who will now be in a position to help the consumer choose the product that best fits the consumer's use patterns.

A related model skips the intermediary and relies on competing sellers as agents of consumers. (On the notion of sellers as agents of consumers, see Gillette 2004.) In the scenario above, for example, the consumer's current carrier is at a competitive advantage because it knows the consumer's use patterns. If the current carrier is required to disclose use information in electronic form, the consumer could then forward this information to competing carriers and ask which of their plans best fits his or her use patterns. This type of

<sup>13</sup> Designing effective TCO disclosures can be a difficult task. First, in many cases the aggregation formula is not obvious and it inevitably involves policy judgments. Second, political economy dynamics may prevent the promulgation of a true TCO disclosure.

disclosure would level the playing field between the old carrier and its competitors, to the benefit of the consumer.

This alternative disclosure paradigm avoids the trade-off between disclosing more information and disclosing information that can be effectively utilized. Since the disclosed information is to be used by sophisticated parties—intermediaries or sellers—rather than directly by consumers, the disclosure can be comprehensive and complex. Disclosure that benefits consumers without being targeted directly at consumers has been prominently proposed by Richard Thaler and Cass Sunstein (2008) (see also Bar-Gill and Board 2012). And the idea is gaining traction among lawmakers around the world (see Sunstein 2010, 2011; chapter 28 by Sunstein in this volume; Federal Communications Commission 2009; Department of Business Innovation and Skills and Cabinet Office Behavioral Insights Team 2011).

### 4.3 Default Rules and Safe Harbors

A second set of policy tools in the soft-paternalism toolkit includes default rules and safe harbors. While mandatory rules say what terms and practices can and cannot be part of a consumer transaction, default rules are more like suggestions—they apply unless the parties opt out. Default rules are becoming increasingly popular. Before the CARD Act, many issuers would allow cardholders to exceed their credit limit, and then impose over-limit fees on these cardholders. The CARD Act set a default rule that prevents cardholders from exceeding the credit limit (under this default, the issuer would decline a charge that takes the cardholder above her credit limit). But it is just a default rule: cardholders can opt out of the default and back into the old regime. A similar default rule has been applied to debit card and ATM transactions: Consumers cannot charge more than the amount available in their checking account, unless they specifically opt into the bank's overdraft protection plan.<sup>14</sup> Indeed, Cass Sunstein, in his prior role as Administrator of the Office of Information and Regulatory Affairs, encouraged the heads of all executive branch agencies to consider the use of default rules as a regulatory tool (Sunstein 2010).

In many contexts default rules can be a powerful instrument. There are several important examples of default rules that affected major changes in behavior and outcomes. When default rules are sticky and few individuals opt out, a default rule can be almost as effective as a mandatory rule, while avoiding the strong paternalism objection. The classic examples include rules that default driver's-license applicants into being organ donors (Johnson and Goldstein 2003) and rules that automatically enroll employees in their employer's 401(k) retirement savings plan (Madrian and Shea 2001). (See also Thaler and Sunstein 2008; Korobkin 1998; chapter 12 by Korobkin in this volume.)

<sup>14</sup> On the new credit card rules—see CARD Act, *supra*. On the debit card and ATM rules—see [http://www.federalreserve.gov/consumerinfo/wyntk\\_overdraft.htm](http://www.federalreserve.gov/consumerinfo/wyntk_overdraft.htm).

There is a concern that default rules would be less effective in the context of consumer transactions. If policymakers set proconsumer defaults, so the argument goes, sellers could easily opt-out in the fine print of their standard form contracts, without allowing the consumer any meaningful opportunity to object (or demand compensation for the opt-out) (see Barr, Mullainathan, and Shafir 2009). While there is valid reason for concern, there is also evidence of effective default rules—defaults that were not subject to wholesale opt-out—in the consumer transaction area. Indeed, the CARD Act rule setting a no-charging-beyond-the-credit-limit default has seen very limited opt-outs.<sup>15</sup>

Moreover, if wholesale, fine-print opt-out is a real concern, policymakers can take steps to increase the stickiness of the proconsumer default. For example, policymakers can require explicit, separate consent, by the consumer, to such opt-out. The opt-out process, and its regulation, provides a range of possibilities for policymakers.<sup>16</sup> (See Ayres 2012; Zamir 1997.)

Safe harbors can be viewed as a type of sticky default. It is not uncommon for the legislator to set a vague standard. The regulator can then step in and define a safe harbor—a course of action that would presumptively satisfy the legislative standard. Firms are not required to use the safe harbor; they are free to adopt other practices. But this de facto opt-out from the de facto default rule, the safe harbor, comes at a price—the price of enhanced regulatory scrutiny (as the practice would need to be evaluated against the vague standard). This enhanced regulatory scrutiny can be viewed as the price of opt-out; it is what makes the default rule sticky. The qualified mortgage, included in the Dodd-Frank Act and defined by the CFPB, offers an example of the safe harbor strategy.<sup>17</sup>

While the default rules described above apply generally to all consumers (unless they opt out), technological innovation opens the door to the design of personalized default rules (see chapter 28 by Sunstein in this volume; Porat and Strahilevitz 2014). Finally, the related tool of active choice should be mentioned. Rather than setting option A or option B as a default, the lawmaker can require the consumer to actively choose between the two options, as a precondition for the consummation of the transaction (see Sunstein 2012).

## 4.4 Right to Withdraw

A third type of soft paternalistic intervention focuses on the consumer's right to withdraw from the transaction. Sometimes referred to as “cooling-off periods,” these rules

<sup>15</sup> See CFPB, CARD Act Factsheet, <http://www.consumerfinance.gov/credit-cards/credit-card-act/feb2011-factsheet/>. See also Johnson et al. (1993) (documenting limited opt-out of default rules specifying insurance coverage).

<sup>16</sup> A sticky default has been proposed, in the mortgage context, by Barr, Mullainathan, and Shafir (2009).

<sup>17</sup> CFPB, Ability to Repay and Qualified Mortgage Standards under the Truth in Lending Act (Regulation Z) (<http://www.consumerfinance.gov/regulations/ability-to-repay-and-qualified-mortgage-standards-under-the-truth-in-lending-act-regulation-z/>).



allow the consumer to renege and cancel the transaction within a specified period of time. Federal Trade Commission rules provide consumers in the United States with a three-day cooling-off period in door-to-door sales.<sup>18</sup> Similarly, Section 125 of the Truth-in-Lending Act imposes a mandatory three-day rescission, cooling-off period for mortgage loans (which is extended to three years if the creditor fails to provide the required disclosure concerning the right to rescind). Many state statutes provide similar protections for particular categories of goods and services, for example, sales of home food service plans, adult and vocational education programs, health studio service contracts, campsite time-shares, and more (see Camerer et al. 2003); and for transactions resulting from certain sales methods, such as sales made by telemarketers.<sup>19</sup> Other jurisdictions similarly mandate withdrawal periods for a growing list of consumer transactions. For example, under Article 9 of the Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011, consumers enjoy a right to withdraw from distance and off-premises contracts, within a 14-day period, at no cost to the consumer.

A right to withdraw does not restrict the range of enforceable contract terms, and is therefore considered to be within the set of soft paternalistic policies. Rather than targeting the terms of the contract, the right to withdraw focuses on the process of contract formation, imposing a delay-period before the contract fully binds the consumer. This forced delay provides the consumer with time to reconsider the costs and benefits of the transaction, free from the pressure of possibly aggressive sales tactics.<sup>20</sup> (See Camerer et al. 2003.)

The right to withdraw also addresses problems of information asymmetry. The ability to inspect the product and try it out for a period of time, before making a final commitment to purchase, is valuable, especially if products are easily returnable. A right to withdraw may also be valuable to the seller, if it increases demand for the seller's products. Consumers would be more likely to make remote purchases if they can return a product that turns out to be less attractive than it initially appeared (see Ben-Shahar and Posner 2011; Eidenmuller 2011).

But, alongside these benefits, a right to withdraw entails potentially large costs, especially when it is abused by a subgroup of opportunistic consumers. Returned items depreciate in value, sometime substantially. This cost will be born, at least in part, by consumers, as sellers anticipate the likelihood of returns and increase prices accordingly.<sup>21</sup> The effect is similar to that of any other mandatory quality feature. Indeed, a

<sup>18</sup> See FTC Rule Concerning Cooling-Off Period Made for Sales at Homes or at Certain Other Locations, 16 C.F.R. pt. 429 (2008). Most states have followed the FTC and enacted similar rules.

<sup>19</sup> See, e.g., Code of Alabama 1975, sec. 8-19A-14.

<sup>20</sup> If the initial purchase decision was made using the quick and intuitive System 1 thought process, then the additional time provided by the right to withdraw allows for the engagement of the deliberative System 2 thought process. See, e.g., Kahneman (2011).

<sup>21</sup> This price increase will hurt poorer consumers. Also, with a mandatory right to withdraw, consumers who are less likely to return the product or cancel the service will be forced to cross-subsidize consumers who are more likely to invoke the right to withdraw. See Bar-Gill and Ben-Shahar (2013).

right to withdraw is a mandatory rule—only one targeting the contracting process, rather than the terms of the contract. Viewed in this light, a right to withdraw can be quite paternalistic.

A less paternalistic alternative would set a right to withdraw as a default rule in certain categories of consumer transactions. Indeed, even in the absence of a mandatory right to withdraw and when the default rule is set to exclude such a right, many prime retailers voluntarily offer a right to withdraw. Even Walmart offers a 90-day free returns policy. This suggests that a default right to withdraw will not suffer from wholesale opt-out. And when opt-out does occur, it may well be efficient—providing a low-price option for consumers who prefer it.<sup>22</sup>

A right to withdraw—be it mandatory, a default that sellers did not opt out of or a voluntary contractual design feature adopted by sellers—can be costly to imperfectly rational consumers who overestimate the benefits of such a right and underestimate the cost of exercising it. The concern is that a consumer, falsely reassured by the right to withdraw, will make a purchase that she will later come to regret. The consumer would then need to incur the cost of withdrawal or, if this cost is too high, keep the undesired good or service (see Becher and Zarsky 2011). Viewed in this light, the right to withdraw is not a more or less paternalistic solution to a behavioral market failure; it is a source of an independent market failure.

## 5 CONCLUSION

---

The structure and design of consumer transactions can be understood as the product of an interaction between market forces and consumer psychology. When consumers make persistent mistakes in evaluating the costs and benefits of a product or service, this interaction results in a market failure—a behavioral market failure—that can substantially hurt consumers and reduce social welfare. When this happens, legal intervention should be considered. While mandatory rules, including bans on certain terms or practices, may sometimes be justified, lawmakers should first try soft paternalistic policies—disclosure regulation, default rules and safe harbors, and a right to withdraw.

This chapter lays out a general theoretical approach to the study of consumer transactions. But there is only so much that one can do at the level of general theory. Market-specific analysis is required to evaluate (1) whether mistakes persist over time, and (2) the welfare costs of consumer mistakes and of sellers' contractual responses to these mistakes. Also, when a persistent, and costly, market failure is identified, market-specific analysis is needed for crafting an appropriate legal policy response.

<sup>22</sup> See Eidenmueller (2011), who identifies the theoretical justifications for withdrawal rights and concludes that they ought to be optional (i.e., default rules) in the distance selling context but mandatory in cases like doorstep sales that involve decision biases that are either preexisting or heightened by sellers.

A final note—on methodology: The preferred method for identifying a behavioral market failure begins with an observed contractual design feature. If the existence (or prevalence) of this feature is difficult to explain under the assumption that consumers are perfectly rational, then the observed design feature itself serves as evidence of consumer misperception.<sup>23</sup> This methodology should be distinguished from the alternative approach—an approach that begins with a cognitive bias, identified in the lab, and proceeds to explore the implications of this bias in the real world. A common critique of this alternative approach questions whether a bias that was identified in a lab experiment would persist in a real-world market setting. This critique is avoided when the starting point is a contractual design feature that is observed in the market.

## ACKNOWLEDGMENTS

---

I thank the editors for their helpful comments and suggestions. The financial support of the D'Agostino/Greenberg Fund at NYU School of Law is gratefully acknowledged. This chapter is based, in part, on prior work, which culminated in Bar-Gill (2012).

## REFERENCES

---

- Armstrong, M. (2008). Interactions between competition and consumer policy. *Competition Policy International* 4:97–147.
- Ayres, I. (2012). Regulating opt-out: An economic theory of altering rules. *Yale Law Journal* 121:2032–116.
- Bakos, Y., F. Marotta-Wurgler, and D. R. Trossen. (2009). Does anyone read the fine print? Testing a law and economics approach to standard form contracts. NYU Law and Economics Research Paper no. 09-40. available at SSRN: <http://ssrn.com/abstract=1443256>.
- Bar-Gill, O. (2004). Seduction by plastic. *Northwestern University Law Review* 98:1373–434.
- Bar-Gill, O. (2012). *Seduction by contract: Law, economics and psychology in consumer markets*. New York: Oxford University Press.
- Bar-Gill, O., and O. Ben-Shahar. (2013). Regulatory techniques in consumer protection: A critique of European consumer contract law. *Common Market Law Review* 50:1–18.
- Bar-Gill, O., and O. Board. (2012). Product use information and the limits of voluntary disclosure. *American Law and Economics Review* 14:235–70.
- Bar-Gill, O., and R. Bubb. (2012). Credit card pricing: The CARD Act and beyond. *Cornell Law Review* 97:967–1018.
- Bar-Gill, O., and K. Davis. (2010). Empty promises. *Southern California Law Review* 84:1–45.
- Bar-Gill, O., and F. Ferrari. (2010). Informing consumers about themselves. *Erasmus Law Review* 3:93–119.

<sup>23</sup> It should be acknowledged that determining whether a certain contractual design feature can or cannot be explained within a rational choice framework may be a difficult task.

- Bar-Gill, O., and R. Stone. (2009). Mobile misperceptions. *Harvard Journal of Law and Technology* 23:49–118.
- Bar-Gill, O., and R. Stone. (2012). Pricing misperceptions: Explaining pricing structure in the cellular service market. *Journal of Empirical Legal Studies* 9:430–56.
- Barr, M. S., S. Mullainathan, and E. Shafir. (2009). The case for behaviorally informed regulation. In *New perspectives on regulation*, D. Moss and J. Cistrenino (eds.). Cambridge: The Tobin Project, 25–61.
- Beales, H., R. Craswell, and S. Salop. (1981). The efficient regulation of consumer information. *Journal of Law and Economics* 24:491–539.
- Becher, S, and T. Zarsky. (2011). Open doors, trap doors, and the law. *Law and Contemporary Problems* 74:63–90.
- Ben-Shahar, O., and E. A. Posner. (2011). The right to withdraw in contract law. *Journal of Legal Studies* 40:115–48.
- Ben-Shahar, O., and C. Schneider. (2014). *More than you wanted to know*. Princeton, NJ: Princeton University Press.
- Camerer, C. et al. (2003). Regulation for conservatives: Behavioral economics and the case for “asymmetric paternalism.” *University of Pennsylvania Law Review* 151:1211–54.
- Craswell, R. (2006). Taking information seriously: Misrepresentation and nondisclosure in contract law and elsewhere. *Virginia Law Review* 92:565–632.
- Della Vigna, S., and U. Malmendier. (2006). Paying not to go to the gym. *American Economic Review* 96:694–719.
- Department of Business Innovation and Skills and Cabinet Office Behavioral Insights Team. (2011). Better choices: better deals, April 13. Available at <http://www.cabinetoffice.gov.uk/resource-library/better-choices-better-deals>.
- Eidenmuller, H. (2011). Why withdrawal rights? *European Review of Contract Law* 7:1–24.
- Ellison, G. (2005). A model of add-on pricing. *Quarterly Journal of Economics* 120:585–637.
- Ellison, G. (2006). Bounded rationality in industrial organization. In *Advances in economics and econometrics: Theory and applications*, Ninth World Congress, vol. 2, R. Blundell, W. K. Newey, and T. Persson (eds.). Cambridge: Cambridge University Press, ch. 5, pp. 142–74.
- Ellison, G., and S. F. Ellison. (2009). Search, obfuscation, and price elasticities on the internet. *Econometrica* 77:427–52.
- Federal Communications Commission (FCC). (2009). Notice of inquiry: Consumer information and disclosure. CG Docket No. 09-158, released August 28.
- Gabaix, X., and D. Laibson. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics* 121:505–40.
- Gilo, D., and A. Porat. (2006). The hidden roles of boilerplate and standard-form contracts: Strategic imposition of transaction costs, segmentation of consumers, and anti-competitive effects. *Michigan Law Review* 104:983–1031.
- Gillette, C. (2004). Rolling contracts as an agency problem. *Wisconsin Law Review* 2004:679–722.
- Glaeser, E. L. (2004). Psychology and the market. *American Economic Review* 94:408–13.
- Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics* 24:461–83.
- Grossman, S. J., and O. D. Hart. (1980). Disclosure laws and takeover bids. *Journal of Finance* 35:323–34.
- Grubb, M. D. (2009). Selling to overconfident consumers. *American Economic Review* 99:1770–807.

- Hillman, R., and J. Rachlinski. (2002). Standard-form contracting in the electronic age. *New York University Law Review* 77:429–95.
- Johnson, E. J., and D. G. Goldstein. (2003). Do defaults save lives? *Science* 302:1338–39.
- Johnson, E. J., J. C. Hershey, J. Meszaros, and H. C. Kunreuther. (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty* 7:35–51.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D., P. Slovic, and A. Tversky (eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kamenica, E., S. Mullainathan, and R. Thaler. 2011. Helping consumers know themselves. *American Economic Review* 101(3):417–22.
- Korobkin, R. (1998). The status quo bias and contract default rules. *Cornell Law Review* 83:608–87.
- Korobkin, R. (2003). Bounded rationality, standard form contracts, and unconscionability. *University of Chicago Law Review* 70:1203–95.
- Loewenstein, G., and T. O'Donoghue. (2004). Animal spirits: Affective and deliberative processes in economic behavior. Available at SSRN: <http://ssrn.com/abstract=539843>.
- Lynch, J. G., and G. Zauberman. (2006). When do you want it? Time, decisions, and public policy. *Journal of Public Policy and Marketing* 25:67–78.
- MacDonald, D. A. (2007). Viewpoint: Card industry questions Congress needs to ask. *American Banker*, March 23, 2007.
- Madrian, B. C., and D. F. Shea. (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics* 116:1149–87.
- Marotta-Wurgler, F. (2007a). What's in a standard form contract? An empirical analysis of software license agreements. *Journal of Empirical Legal Studies* 4:677–713.
- Marotta-Wurgler, F. (2007b). Unfair dispute resolution clauses: much ado about nothing? In *Boilerplate: foundations of market contracts*, O. Ben-Shahar (ed.). Cambridge: Cambridge University Press, pp. 45–65.
- Marotta-Wurgler, F. (2011). Does increased disclosure help? Evaluating the recommendations of the ALI's "Principles of the Law of Software Contracts." *University of Chicago Law Review* 78:165–86.
- McClure, S., D. Laibson, G. Loewenstein, and J. Cohen. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science* 306:503–7.
- Milgrom, P. (2008). What the seller won't tell you: Persuasion and disclosure in markets. *Journal of Economic Perspectives* 22:115–31.
- Nalebuff, B., and I. Ayres. (2003). *Why not?* Boston: Harvard Business School Press.
- O'Donoghue, T., and M. Rabin. (1999). Doing it now or later. *American Economic Review* 89:103–24.
- Porat, A., and L. J. Strahilevitz. (2014). Personalizing Default Rules and disclosure with Big Data. *Michigan Law Review* 112:1417–1478.
- Radin, M. J. (2012). *Boilerplate: The fine print, vanishing rights, and the rule of law*. Princeton, NJ: Princeton University Press.
- Sovern, J. (2006). Toward a new model of consumer protection. *William and Mary Law Review* 47:1635–709.
- Spiegler, R. (2006). Competition over agents with boundedly rational expectations. *Theoretical Economics* 1:207–31.
- Spiegler, R. (2011). *Bounded rationality and industrial organization*. New York: Oxford University Press.

- Sunstein, C. R. (2010). Disclosure and simplification as regulatory tools, memorandum for the heads of executive departments and agencies. Available at [http://www.whitehouse.gov/sites/default/files/omb/assets/inforeg/disclosure\\_principles.pdf](http://www.whitehouse.gov/sites/default/files/omb/assets/inforeg/disclosure_principles.pdf).
- Sunstein, C. R. (2011). Informing consumers through smart disclosure, memorandum for the heads of executive departments and agencies. Available at <http://www.whitehouse.gov/sites/default/files/omb/inforeg/for-agencies/informing-consumers-through-smart-disclosure.pdf>.
- Sunstein, C. R. (2012). Impersonal default rules vs. active choices vs. personalized default rules: A triptych. Available at SSRN: <http://ssrn.com/abstract=2171343> or <http://dx.doi.org/10.2139/ssrn.217134>.
- Sunstein, C. R. (2013). The Storrs Lectures: Behavioral economics and paternalism. *Yale Law Journal* 122:1826–99.
- Sunstein, C. R., and R. H. Thaler. (2003). Libertarian paternalism is not an oxymoron. *University of Chicago Law Review* 70:1159–202.
- Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica* 47:143–48.
- Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making* 12:183–206.
- Thaler, R. H., and Sunstein, C. R. (2003). Libertarian paternalism. *American Economic Review* 92(2):175–79.
- Thaler, R. H., and Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth and happiness*. New Haven: Yale University Press.
- Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality & Social Psychology* 39:806–20.
- Zamir, E. (1997). The inverted hierarchy of contract interpretation and supplementation. *Columbia Law Review* 97:1710–803.

## CHAPTER 19

---

# BEHAVIORAL ECONOMICS AND INSURANCE LAW

### *The Importance of Equilibrium Analysis*

---

TOM BAKER AND PETER SIEGELMAN

## 1 INTRODUCTION

---

THE discipline of economics and the field of insurance have had a long, mutually productive encounter. The economics of information grew out of ideas first articulated by insurance actuaries and then formalized by economists, some of whose life paths had taken them through the insurance business (Baker 1996; Baker 2003). Mathematicians hired by insurance organizations in the eighteenth and nineteenth centuries developed statistical techniques that, with later advances, led to the subfield of econometrics (Daston 1988; Clark 1999; Alborn 2009). More recently, insurance institutions have employed insights and techniques from many parts of the economics discipline: insurance economics to be sure, but also financial economics, health economics, econometrics, and, last but not least, law and economics. One measure of the insurance industry's appreciation of economics can be seen in the International Association for the Study of Insurance Economics (the Geneva Association), the members of which are the chief executive officers of 80 insurance companies worldwide, and which sponsors the journal *Geneva Papers on Risk and Insurance*, which has published work by many leading economists. Law and economics is a dominant paradigm in insurance legal scholarship (Abraham 1986), and has had an impact on the development of insurance law, among other ways through the prolific insurance law opinions of Judge Richard Posner (Langer 2001; ALI Liability Insurance Project Preliminary Draft No. 1 2013).

The US Affordable Care Act (§ 1501(a)(2)(I), as amended by § 10106; National Association of Insurance Commissioners, Exchanges (B) Subgroup, 2011) provides a striking example of the influence of economic ideas and economists on the insurance field, as well as the feedback loop between insurance practice and economic theory. The



Obama administration justified the most politically and legally controversial aspect of the act (the mandate that individuals buy health insurance) on the basis of an economic idea (adverse selection) that is directly traceable to the encounter between insurance practice and economic theory.

[I]f there were no requirement [to maintain minimum essential coverage], many individuals would wait to purchase health insurance until they needed care. By significantly increasing health insurance coverage, the requirement, together with the other provisions of this Act, will minimize this adverse selection and broaden the health insurance risk pool to include healthy individuals, which will lower health insurance premiums. The requirement is essential to creating effective health insurance markets in which improved health insurance products that are guaranteed issue and do not exclude coverage of preexisting conditions can be sold. (Affordable Care Act, at § 1501(a)(2)(I), as amended by § 10106)

The recent behavioral turn in economics provides another opportunity to learn from this encounter. Insurance provides a fertile testing ground for, and potential challenge to, standard economic theory. The product at issue is relatively straightforward: contingent claims on money. Thus, theory generates clear testable predictions and normative statements about when rational people will and should buy what kinds of insurance, provided of course that the institutional context can be adequately specified. Empirical and experimental research reveals consistent, reproducible patterns of behavior that depart from these predictions, however. This divergence between clear theoretical predictions and empirical findings poses a series of challenges—to the adequacy of the specifications of the institutional context underlying the predictions, to the rationality of the observed behavior, to the regulatory framework that shapes the insurance market, and ultimately to standard economic theory itself.

In this chapter we report on and engage with this ongoing, productive, and sometimes frictive encounter. As law professors with a substantial investment in understanding insurance institutions, we are especially interested in charting (and influencing) the meaning of this encounter for insurance regulation. But the payoffs from this exercise, like others in the long-term relationship between insurance and economics, extend beyond the specific problem at hand.

After first reviewing some basic economics of insurance and behavioral research, we closely analyze two types of insurance. Although both types are widely purchased, neither should be appealing to most rational, reasonably informed individuals, because the net benefits they provide are almost certainly negative. The first—extended warranties for consumer products—is among the most profitable forms of insurance. This suggests both that the *perceived* benefits to purchasers very substantially exceed what expected utility theory would predict and that the market is not working to deliver the insurance protection at a reasonable price (Baker and Siegelman 2013). Our second case study—low deductible homeowners' insurance—looks similar to the first in some ways. Homeowners' insurance itself is valuable to a rational actor, but most people buy policies with deductibles that are far too low to be justified in expected utility terms.

But there are some subtle, yet very significant differences between these two markets, as we demonstrate below.

These may seem like odd and perhaps even inconsequential phenomena upon which to direct serious analytical firepower. Yet they are ideal for the task at hand. First, the observed behavior sharply diverges from the predictions of expected utility theory. Second, the institutional context is sufficiently well understood that we can be reasonably sure that the divergence reflects something about consumer behavior, not the difficulty of developing sound predictions for that context. Third, the behavioral research is sufficiently developed to provide a well-grounded explanation (or set of explanations) for consumer behavior and why it diverges from the predictions of expected utility theory. Finally, as others have already pointed out (Schwarcz 2010; Camerer et al. 2003; Braun and Muermann 2004), this research might well point toward a sophisticated, “consumer sovereignty” justification for kinds of insurance that expected utility theory would condemn, posing a clear challenge to that theory. Both examples thus present difficult, practical, and generalizable problems for regulators: how should policymakers respond when consumers apparently “want” (or at least, are willing to pay for) something that a rational person would not choose to buy?<sup>1</sup> What, if any, forms of regulation are likely to be effective and desirable under these conditions?

## 2 THE ECONOMICS OF INSURANCE IN BRIEF

The discipline of economics has a simple but powerful explanation of the value of insurance to individuals, a well-worked-out explanation of why insurance needs to be regulated, and a relatively consistent approach to the form that insurance regulation should take.

In this paradigm, people value insurance for two reasons: they are risk averse (meaning that they have a declining marginal utility of money) and, with regard to contingent

<sup>1</sup> By way of contrast, consider *United States of America v. Rose Marks et al.*, Case No. 11-80072-CR-MARRA/VITUNAC(s) (S.D. Fl. 2011) (available at <http://www.justice.gov/usao/fls/PressReleases/Attachments/110816-01.SupersedingIndictment.pdf>), in which the defendant and nine others are charged, among other things, with mail and wire fraud, for claiming “to have powers of intuition that enabled [them] to perceive things beyond the realm of the five senses,” and “represent[ing] to [their] clients that [they were] conferring with the Archangel Michael for his advice and counsel for them.” One defendant was accused of having told a client that “they would bring her [estranged] husband back to her, . . . [but that this] ‘work’ would require sacrifices which would mean money because money was the root of all evil.” *Id.* at 12. The US Attorney alleges that the amount wrongfully taken from clients exceeded \$40 million. See <http://www.justice.gov/usao/fls/PressReleases/110816-02.html>. Most readers presumably wouldn’t question such indictments, although extreme devotees of caveat emptor might believe that fraud should not be criminalized. But surely the patrons of the alleged psychics believed that they were getting something worthwhile for the money they spent, just as buyers of extended warranties do. Both psychics and extended warranties can and do make people feel better; both do so only by appealing to their irrational natures.

losses, insurance is a more efficient way than savings to equalize the marginal utility of consumption over time. Put in ordinary language, insurance allows people to shift money from times when they do not need it very much to times when they need it much more. Expected utility theory teaches that insurance is most valuable when it provides a mechanism for a large group of people to each pay a small amount of money so that there is a large sum available for the few who really need it. The do-it-yourself alternative—savings—is not as efficient, because it shifts the money into the future whether you need it then or not. By contrast, insurance gives you the money in the future only if you need it.

In a world with perfect information and no transaction costs, every risk-averse person would be better off with insurance against all risks, actuarially fair insurance (with a premium just equal to the expected loss) would be available for all risks, and there would be no need for regulation of insurance markets (Arrow 1971a). (Risk-neutral or risk-seeking people would not demand insurance even in this world; but there is very little evidence for the existence of risk-seeking or risk-neutral preferences.) Even if risks are correlated rather than independent of each other, this proposition is still essentially correct (Jaffee 2006). Adding a dose of realism by acknowledging the presence of transaction costs (i.e., the costs of selling the insurance and running the insurance business) changes this conclusion only slightly: instead of complete insurance, people would be better off with partial insurance, such as insurance with a deductible or coinsurance. With perfect information, once again, the market would supply the appropriately partial form of insurance, with no need for regulation.

The need (and economic justification) for insurance regulation becomes apparent after adding a second, larger dose of realism: taking into account the information problems that exist on both sides of the insurance relationship.

## 2.1 Information Problems Facing Insurers

Insurance companies have long been aware of the information problems that exist on their side of the relationship: there are limits to what an insurance company can find out about the people looking for insurance, and it's hard to monitor consumers' behavior once they have it. The result is that the people buying insurance tend to be more risky than average (the adverse selection problem) and, once they have it, they aren't as careful to avoid losses as they would have been without the insurance (the moral hazard problem).

Economists have formalized the insurance companies' information problems and, in the process, developed what has come to be known as the economics of information. Insurance purchasers' private information about their risk leads to adverse selection, the information problem that George Akerloff (1970) first discussed in his Nobel Prize-winning paper on the "lemons problem" (Rothschild and Stiglitz 1976, and for a recent empirical survey, Cohen and Siegelman 2010). The insurer's inability to monitor its customers' behavior after they buy insurance leads to moral hazard, the

information problem that Kenneth Arrow (1963)—another Nobel Prize winner and a former insurance actuary—first discussed in his classic article on the economics of health insurance.<sup>2</sup>

Insurance companies' information problems offer parsimonious, powerful explanations of much of the institutional structure of real-life insurance relationships. As economic theory would predict, and as evidence confirms, insurers manage moral hazard through several devices. First, cost-sharing arrangements such as deductibles or copays give policyholders some "skin in the game" by leaving them with some of the potential losses that their conduct might cause. Second, insurance contracts expressly exclude coverage for certain kinds of losses (such as those caused by a policyholder's reckless or deliberate behavior) (Shavell 1979). This, again, limits policyholders' incentives to slack off on precautions. Insurers will often engage in precontract underwriting (screening) that attempts to discern the honesty, prudence, and trustworthiness of insureds and denies coverage to those who meet certain minimum requirements. Finally, insurers rely on social norms to prevent excessive slacking-off in precautions (Heimer 1985; Arrow 1963)

Insurers manage adverse selection through an array of similar devices. Risk classification entails the use of verifiable measures (such as smoking status or age or historical loss data) that correlate with risk to set premiums. Another technique relies on contract terms that encourage long-term relationships (Hendel and Lizzeri 2003). Providing a menu of insurance policies that induces policyholders to sort themselves according to their riskiness is another approach insurers use to curtail adverse selection (Rothschild and Stiglitz 1976). Underwriting can also be useful in controlling adverse selection, since insureds' informational advantage diminishes as the insurer learns more about them. For the most part, insurance organizations have been able to arrive at reasonably satisfactory solutions to these problems without the government's help. Exceptions to privately arranged solutions include the mandated purchase of insurance to prevent adverse selection (as in the Affordable Care Act) and the related regulation of competition among insurers to prevent cream-skimming (attempts to attract only the best risks) and other behavior that, in the limit, can sometimes lead to the failure of

<sup>2</sup> Arrow 1963. Arrow worked as an insurance actuary before going to economics graduate school. He summarized the impact of his actuarial experience as follows:

one thing is true ["about this actuary business"]: I really learned. One thing I learnt about during the course of this was moral hazard and adverse selection. . . . I suddenly realized insurance people knew what they were talking about: there was a real economic issue which economists had not understood. It turned out that even though I didn't pursue it [at the time], it was a very important economic problem. I really understood what risk bearing was about and understood the realities of it.

Interview with Kenneth Arrow by Juan Dubra. Munich Personal RePEc Archive, March 2005. Available at <http://mpra.ub.uni-muenchen.de/967/>. Joseph Stiglitz, another Nobel laureate who worked on insurance information problems, was the son of an insurance agent (personal communication).

insurance markets (Cutler and Zeckhauser 1998; but see Siegelman 2004 for the relative infrequency of insurance market failures).

## 2.2 Information Problems Facing Consumers

Insurance regulators have long been aware of the information problems that exist on the consumer side of the insurance relationship. Left on their own, ordinary consumers can know very little about either the insurance they're buying or the companies selling that insurance; and once they buy insurance, they are vulnerable to insurer opportunism because they cannot observe, for example, whether an insurer provides satisfactory claims-handling service (Sulzle and Wambach 2005). These information problems are less susceptible to private contract-based solutions than those facing insurers, and thus provide wider ranging justification for insurance regulation. (Brokers or middlemen might in theory solve some of these problems, and are in fact commonly observed in insurance markets. But brokers create problems of their own, and in any case, are typically unavailable for smaller-scale transactions.) Insurance companies have private information about many things that affect the value of their products: for example, their solvency (a promise to pay is not worth much if the company is not able to pay), the meaning of the terms of their contracts, and their approach to investigating and paying claims.

Referring to this information as "private" does not mean that it is completely unobservable. For example, the written terms of an insurance contract appear in the insurance policy form (assuming that the insurance company is willing to provide the policy in advance, which is not always the case in practice) (Schwarcz 2011). But it is so time-consuming and expensive to evaluate the terms of the contract or, indeed, most of the other observable aspects of quality, that no individual person or company would rationally make that effort (Harel and Procaccia 2009). Other aspects of quality, such as past claims-servicing practices or current financial solidity, might be observable in theory, but that observation would require the disclosure of information that the insurer prefers to keep private and that is interpretable only in relation to information about other insurers (posing a collective action problem, a classic justification for regulation). Still other aspects of insurance product quality are completely unobservable by anyone at the time of purchase, because they depend on what happens in the future. Insurance consists fundamentally of the promise to pay money in the future, sometimes very far in the future. No one can observe today the financial solidity and claims-paying practices of an insurance company in the future.

This private information creates the potential for (inverse) adverse selection, the risk that bad insurance contracts will drive out the good, and (inverse) moral hazard, the risk that insurance companies will change their financial condition and claims-paying practices to the detriment of existing policyholders (Beal 2000–2001). Insurance regulation addresses these problems by certifying the quality of both insurance contracts and insurance companies. Government approval of insurance companies' standard

form contracts certifies the quality of those contracts. Solvency regulation and insurance guarantee funds certify the insurance companies' ability to pay claims. Market conduct regulation and related tools such as private rights of action for insurer misconduct, in effect, certify insurance companies' willingness to pay claims and deter insurers from opportunism at the point of claim.

### 2.3 Insurance Economics and Insurance Regulation

The economic approach to insurance has been enormously influential among people who study and teach insurance, in actuarial training and practice, and, as a result, within at least the expert sector of the insurance regulatory community. Much of insurance regulation is broadly consistent with the economics of insurance, even if the actual implementation of regulation may fall short of economic prescriptions (Chandler, n.d.)

Nevertheless, our sense is that the standard economic model does not easily justify all of the consumer protection rationales for insurance regulation. The reason is that, in the standard economic model, consumers—bolstered by competition among insurers—are assumed to be reasonably well equipped to maximize their own utility, so that intervention by insurance regulators is likely to deprive consumers of choices they would either prefer to make (in which case, the consumers experience a loss of welfare) or would not make (in which case, the regulatory intervention is simply useless). The behaviorally informed research that we review next presents a very different view of consumer behavior, in which consumers are poorly equipped to maximize their own utility, and of markets, in which firms are able to avoid the leveling effect of competition.

## 3 BEHAVIORAL ECONOMICS AND INSURANCE

---

Much as the insurance market has provided fertile ground for the development of the economics of information, it has also spurred the growth of behavioral economics. Researchers such as Howard Kunreuther (Johnson et al. 1993; Shoemaker and Kunreuther 1979; and for a recent comprehensive treatment, Kunreuther, Pauly, and McMorrow 2013) and others (Cutler and Zeckhauser 2004) have long noted that consumer behavior in the insurance market does not match the predictions of standard economic theory: anomalies abound. Consumers do not demand enough of some kinds of insurance that the standard account says that they should value highly, such as insurance against large-but-infrequent catastrophes (earthquakes, floods) and annuities (longevity insurance). At the same time, consumers demand too much of some other kinds of insurance that, in theory, they should not want at all, such as dread disease insurance and extended warranties for consumer products. And, given the

choice, consumers regularly purchase insurance policies with deductibles and policy limits that are too low relative to the costs and benefits (Kunreuther, Pauly, and McMorro 2013; Cutler and Zeckhauser 2004). For example, Martin Feldstein (1973) has famously suggested that for reasonable levels of risk aversion, optimal health insurance would entail a much greater level of risk-sharing than current health insurance, with co-insurance rates of 50%–66%, meaning that the policyholder is responsible for one-third to one-half of any insured medical expense. (Although risk-averse consumers would ideally choose to insure 100% of all exogenous risks when prices are actuarially fair, full insurance is no longer optimal when insurance is subject to moral hazard or overuse that increases premiums. In Feldstein's [1973, 251] analysis, reducing insurance coverage by one-third would lead to an increase in welfare equivalent to about 25% of total private insurance premiums.)

### 3.1 Demand-Side Anomalies

Research has revealed an increasingly well-defined set of what Kunreuther and Pauly call “demand side anomalies” in the insurance market, and we draw heavily on their taxonomy in what follows (Kunreuther, Pauly, and McMorro 2013). These anomalies are regularities in insurance purchasing behavior that differ systematically from what expected utility theory predicts. A great deal is at stake in the use of the word “systematically,” as we demonstrate below. Even if replicable laboratory experiments can isolate particular biases in highly controlled environments, behavioral research often lacks a meta-theory about which biases will be operative in complex real-world settings (Barberis 2013). As we suggest, this is especially significant for regulators, because biases seem to be context-dependent and of uncertain signs—that is, some biases lead to “too much” insurance being purchased, while others lead to too little. The following is a selective list:

- People choose low deductibles and, because of loading costs, overpay to provide protection against losses that are not worth insuring against, given plausible levels of risk aversion (Sydnor 2010).
- Having chosen and paid for a low deductible, people do not file a claim unless their loss is much larger than the deductible. One study found that 80% of households had a pseudodeductible higher than the next highest available deductible, meaning that they could save money without affecting their actual coverage by selecting the higher deductible policy (Braun et al. 2006).
- People buy some kinds of insurance that protect exclusively against losses that are small in relation to their wealth, sometimes even when the price for that insurance is quite high in relation to its expected value. For example, many people buy extended warranties for consumer durables such as TV sets, or purchase collision damage waivers for rental cars (coverage for losses to the rented vehicle only). Such purchases are per se irrational under the standard theory of risk aversion,



which applies *only* to losses that are a large fraction of one's total wealth (Hogarth and Kunreuther 1995; Karla et al. 2009; Rabin and Thaler 2001; Jindall 2013; Baker and Siegelman 2013).

- People are more willing to insure emotionally treasured objects than they are to insure other objects of equal financial value, and they also put more effort into preparing insurance claims for the loss of a treasured object (Hsee and Kunreuther 2000). The standard economic model of insurance demand is predicated on risk aversion, which implies a decreasing marginal value of money. On this account, subjective value—for example, for family heirlooms—should not motivate insurance purchases unless the loss of the object would increase the marginal utility of wealth. Thus, it is irrational to insure grandpa's shaving mug (market value \$50) unless losing the mug would make an additional dollar (substantially) more valuable than if the mug were intact.
- People prefer insurance policies with no-claim rebates or deferred dividends (that is, policies that return some of the policyholder's premiums in the event that no claims are made on the policy), even though these policies violate the assumption of declining marginal utility of wealth (Slovic et al. 1977; Baker and Siegelman 2010).
- Insurance against “named events” (limited purpose life insurance policy in the form of flight insurance, or “dread disease” insurance) is sometimes more attractive than objectively more valuable general insurance (Kunreuther and Pauly 2005).
- People do not buy objectively valuable insurance against other low-probability, high-severity events (Camerer and Kunreuther 1989; Krantz and Kunreuther 2007).
- People are more likely to buy disaster insurance after a disaster, even when they (wrongly) believe that this disaster *reduced* the likelihood of the next one (Kunreuther et al. 1985).

We can broadly group the behavioral explanations for these anomalies into two categories. The first set of explanations focuses on biases that affect the perception of the value of insurance in a manner that conflicts with expected utility theory. Some of these biases tend to decrease the perceived value of insurance and, thus, may lead to insufficient demand. These include the following:

- Excessive discounting (an irrationally high preference for money today over money tomorrow)
- Overoptimism bias (believing that bad things are unlikely to occur to one's self (chapter 13 by Williams in this volume). Note that at least in some formulations, optimism bias may represent an even deeper form of irrationality than is contemplated in most behavioral models, since it may be “inconsistent with the independence of decision weights [e.g., probabilities] and payoffs found in models of choice under risk, such as expected utility, subjective expected utility, and prospect theory” (Bracha and Brown 2010).

Other biases tend to increase the perceived value of insurance and, thus, may lead to exaggerated demand, including the following:

- Loss aversion (the marginal disutility of loss exceeds the marginal utility of gain). Although paying for insurance is an out-of-pocket expense, many consumers frame this as a “cost” or “price,” while perceiving an uninsured loss as a true “loss” (Johnson et al. 1993).
- Emotional attachment to people or objects, which should not influence insurance demand, unless the loss changes the marginal utility of wealth (Hsee and Kunreuther).
- Superstition (buying insurance in the belief that it will prevent bad things from happening).

Still others could have either effect, depending on context, including the following:

- The availability heuristic (risks that are easier to recall are assumed to be more likely to occur than they actually are, and vice versa) (Tversky and Kahneman 1982; Keller et al. 2006)
- Regret aversion (wanting to have made the optimal choice, as determined ex post (Loomes and Sugden 1982). Adding regret aversion to a model of insurance demand leads individuals to “hedge their bets” by purchasing more insurance for small losses and less insurance for large losses than would be optimal from an expected utility perspective (Braun and Muermann 2004).
- Threshold effects (ignoring probabilities below a cutoff in some situations, and, in others, overweighting reductions from an extremely low probability to a perceived zero probability) (Krantz and Kunreuther 2007)
- Overconfidence (sometimes called the control illusion). This bias reduces the perceived likelihood (or effect) of events you can control (car crash) and increases the perceived likelihood (or severity) of events you can’t control (plane crash), making you less likely to buy car insurance and more likely to buy flight insurance.
- Herding (copying friends and family)

Most of the demand-side anomalies listed earlier can be explained by some combination of these biases. Regret aversion helps explain buying insurance for low-value losses (if a loss happens, I don’t want to regret not having the insurance) and buying insurance with a no-claim rebate (if the risk doesn’t materialize, I can be sure I get at least something for my money (Johnson et al. 1993; Baker and Siegelman 2010). So, too, loss aversion and mental accounting: the prospect of future “loss” weighs more heavily than the small additional “price” paid to buy a lower deductible, an extended warranty, or any other low-value insurance sold in connection with another product or service. Emotional attachment helps explain buying insurance for treasured objects. The availability heuristic and dread help explain buying insurance for named events. Threshold

effects help explain not buying insurance for low-probability, high-severity events. The availability heuristic helps explain buying that same insurance after a disaster.

A second set of explanations for the demand-side anomalies focuses on more general information-processing problems that consumers face in making decisions of all kinds. These kinds of behavioral regularities are different from those listed earlier because they don't directly affect the perceived value of insurance. Rather, they reduce the capacity to make a decision, whatever the perceived value of insurance may be. These include the following:

- Hyperbolic discounting (valuations that fall rapidly for small delays, but more slowly for longer delays, leading to procrastination and other time-inconsistent preferences)
- Complexity aversion (avoidance of options that are complicated to evaluate) (Bruce and Johnson 1996)
- Aversion to contemplating some topics (death, stigmatized or taboo events) (Chan 2012)
- More general cognitive constraints. Fredrick (2005) finds, for example, that a simple three-item "Cognitive Reflection Test" can predict such aspects of individual behavior as risk preferences and time preferences, and speculates that "some preferences are better than others and that cognitive ability is one indicator of the 'better' preference. Dohmen and coauthors (2010) conclude that individuals with higher cognitive ability are more willing to take risks and are more patient than those with lower cognitive ability.

In the insurance context, these information-processing problems can lead consumers to make the default "decision" not to buy insurance, or leave them vulnerable to firms that frame or create a bad decision as the default. For example, consumers are often vulnerable to high-pressure sales tactics that encourage them to buy extended warranties on consumer durable items, even when such insurance is massively overpriced by any measure, and even when a rational consumer would not choose to insure such relatively small losses in the first place. Segal (2012) provides anecdotal evidence of abusive seller practices; UK Competition Commission (2003) offers more systematic details.

Considering all of these biases and information-processing problems together produces a rather bleak picture, at least for those who would like to see behavioral economics provide clear guidance to policymakers. There are systematic yet conflicting biases that affect the perceived value of insurance. Consumers want too much of some "bad" kinds of insurance and not enough of some "good" kinds of insurance, and, even if they are motivated to distinguish between good and bad insurance, information-processing problems make doing so very difficult.

In a world of complete information and zero transactions costs, actuarially fair insurance is always and everywhere a valuable financial product for a rational, risk-averse consumer. In the real world, insurance is only sometimes a good financial

deal. Whether it is a good deal in any particular situation is a complicated question that turns on individual preferences, and the frequency and severity of loss, and the loading charges that insurance companies must impose in order to run their business, not to mention the complications resulting from moral hazard, adverse selection, and the existence of alternative ways to manage risk. The behavioral decision research clearly demonstrates that people do a remarkably poor job at making decisions that involve even simple mathematical concepts, such as the compounding of interest (Kunreuther, Pauly, and McMorro 2013). Insurance is a much more complicated financial product than a bank account or loan (Jackson 1999), so it should come as no surprise to learn that behavioral decision research provides very little reason to be confident that consumers are making optimal insurance-purchasing decisions.

### 3.2 Protecting the Imperfectly Rational

Suppose we take it as a given that consumers cannot be relied upon to make wise choices with respect to insurance: What role does this then leave for policy interventions to improve welfare? Our message here is that even if we know the causes and direction of consumer “errors,” the behaviorist turn makes good regulation of insurance more, or at least no less, difficult than it ever was. In this section we discuss some general problems with the design of regulation to protect imperfectly rational insurance buyers.

A key problem for regulators seeking to act on behavioral insights is that behavioral theories may do a good job of *explaining* behavior, but they do so in a way that severs the connection between a consumer’s behavior and her welfare. Under standard economic assumptions, there is a tight link between the two: behavior is chosen to maximize welfare—indeed, this is close to the very definition of rationality. It would be irrational if one preferred X to Y, yet chose Y when both options were possible. A rational consumer who chooses to buy an extended warranty is *by definition* doing so because she believes it advances her welfare, and there’s an obvious subjective sense in which she must be right. The link between behavior and welfare is what gives economics much of its normative bite: allowing consumers to act as they choose is desirable precisely because their actions will be rationally chosen to promote their well-being.

But what happens if the link between behavior and welfare is broken or attenuated, which is precisely the conclusion of the behavioral research in insurance? One obvious regulatory solution—which unfortunately rarely works—is to provide the consumer with the correct information about the relevant risks involved. Doing so poses relatively few problems for a preference-based, welfarist approach to policymaking: by assumption, the consumer will use the new, accurate information to make the appropriate (subjectively welfare-maximizing) choice not to buy the insurance. If on the other hand, providing the information does *not* alter the consumer’s decision, then (arguably) buying the insurance must have been based on some kind of nonstandard preferences (for example, regret aversion or loss aversion), and therefore the purchase

actually increases the consumer's utility. Either way, disclosure appears to solve the problem.

Camerer and coauthors (2003, 1253–54) and Schwarcz (2010) follow this line of reasoning in arguing that mistakes can and should be corrected by disclosure, but that if consumers are (irrationally) buying extended warranties because of loss aversion or as relief for “anxiety,” they should be free to do so, because restricting their ability to make such decisions would leave them (subjectively) worse off. Schwarcz, for example, writes that behavioral anomalies in the purchase of insurance

can plausibly be explained as sophisticated consumer behavior to manage emotions such as anxiety, regret, and loss aversion. Moreover, the capacity of insurance to address these negative emotions is not necessarily an artifact of manipulative insurance sales or marketing. Rather, it may be a sophisticated and informed strategy on the part of consumers to manage emotions that exist independently of insurers' (and their agents') sales efforts.

Yet behavioral (and other) research has not been kind to the proposition that disclosure corrects decisional errors (Ben-Shahar and Schneider 2011; Willis 2008). Precisely because consumers who buy extended warranties are not fully rational, frequency-of-repair statistics and other forms of “debiasing” education will be difficult for them to process. Behavioral research might help to make disclosure more effective, but we see no reason to be optimistic that disclosure can fully overcome even the most minimal behavioral impediments to appropriate decision-making. This in turn implies that the distinction between mistakes (based on incorrect information) and nonstandard preferences as motives for insurance purchases does not provide a solid basis for regulatory policy. Unless we define “mistakes” tautologically (as those decisions that can be altered by disclosure), *effectively* correcting mistakes will often require something more than disclosure, and thus entails making it difficult or impossible for consumers to do what they “want.”

A second important problem—typically only implicit in much of the behavioral research—is the possibility of heterogeneity among consumers. While rational consumers are all alike (in their rationality, if not their preferences), there are a multitude of ways to be irrational. Not only are some people apparently “more rational” than others (Choi et al. [2011] find considerable heterogeneity among subjects, with richer and better educated subjects more likely to exhibit rational behavior); the multiplicity of possible irrationalities adds enormous complexity to policymaking because it means that the conventional Kaldor/Hicks or Pareto criteria for policy evaluation are often unavailable, and distributional issues cannot be avoided.<sup>3</sup> Policies that help one group

<sup>3</sup> We do not mean to suggest that distributional concerns *should* be avoided, but only that heterogeneity makes it impossible to ignore these issues. Choi (2011), Fredrick (2005), and Dohmen and coauthors (2010) all suggest that there are nonsurprising correlations between class, gender and ethnicity, and financial “mistakes,” which in our view only strengthens the case for regulatory intervention.

of irrational consumers may hurt another. Policies that help the rational may harm the irrational, and vice versa. We have relatively little to offer here, except to say that behavioral heterogeneity makes policymaking even more difficult than it would be in a world where consumers were all fully rational.

Finally, when considering whether to modify insurance law or regulation to take the behavioral economic findings into account, it is vital to acknowledge the potential effects of intervention in equilibrium, after all relevant actors have had a chance to adjust their behavior (Schwartz and Wilde 1979). At least since the pioneering work of Rothschild and Stiglitz (1976), economists have understood that equilibrium in insurance markets—which are pervasively characterized by asymmetric information (as described earlier)—can be extraordinarily complex and in some cases might not even exist at all. Adding behavioral “anomalies” to the equilibrium analysis is far from straightforward. But without such an analysis, regulatory interventions are likely to have unintended consequences, and may even be welfare-reducing.

Consider the possibility of overoptimism—consumers’ mistaken belief that they have a lower risk of some loss than is actually the case. One might naturally conclude that this cognitive bias would lead to an inappropriately low demand for insurance, and thus result in welfare losses from excess exposure to risk. And one might be tempted to conclude that education or “nudges” should be deployed to give consumers a more appropriate sense of the risks they face. But in an equilibrium model with asymmetric information, that conclusion no longer holds. In an elegant paper, Sandroni and Squintani (2007) show theoretically that overoptimism can actually *improve* welfare in the presence of adverse selection. That is, when some high-risk insureds optimistically (but mistakenly) believe that they are low risk, they are less inclined to purchase insurance than they would otherwise be. That makes selection problems less severe, and the market actually reaches a better equilibrium as a result. So efforts to debias consumers by giving them a better sense of the probability of loss can correct one problem (overconfidence) only to exacerbate another (selection), in way that might well be welfare-reducing. (On the challenges facing behaviorally informed regulation, see also chapter 28 by Sunstein; chapter 6 by Pi, Parisi, and Luppi; and chapter 7 by Mitchell in this volume.)

The moral of these examples is *not* that behavioral economics offers little or no scope for welfare-enhancing intervention. It is rather that when nonstandard motivations or imperfect reasoning combine with informational asymmetries, policy interventions need to be very carefully tailored to particular circumstances in order to be effective; there is no simple route from identifying a behavioral flaw (itself a complicated endeavor) to recommending an appropriate regulatory policy. We illustrate this conclusion below, in our analysis of low deductible homeowners’ insurance and extended warranties. Both forms of insurance are a bad buy in expected utility terms, but an equilibrium analysis suggests very different regulatory responses.

## 4 TWO EXAMPLES

---

In this section, we offer a more detailed behavioral analysis of two kinds of insurance, extended warranties for consumer products and low-deductible homeowners insurance. Both markets reveal significant anomalies: many consumers make choices that are essentially incompatible with rational behavior. And yet the equilibrium that results, and the scope for policy interventions, differs widely across these two ostensibly similar situations.

### 4.1 Extended Warranties for Consumer Products

Extended warranties (EWs) are optional insurance contracts that are sold with many products to supplement the standard warranty provided by the manufacturer (they are entirely distinct from ordinary warranties that are provided by the manufacturer and that serve as a signal of product quality; see Spence 1977; Priest 1981). Since they are essentially unregulated, it is difficult to know the dollar value of EWs sold each year, but estimates are in the range of \$16 billion for the United States (Baker and Siegelman 2013).

Extended warranties are the quintessential bad deal (at least for a rational, expected utility maximizer) for two related reasons. First, a rational person should not demand insurance for losses that are small relative to his or her wealth, even if he or she is risk-averse. As several distinguished economists have noted, classical risk-aversion stems from—indeed, it just *is*—the declining marginal utility of wealth. But a small loss, by definition, doesn't change wealth very much, and so it simply can't change the marginal utility of wealth by much either (Arrow 1971b; Pratt 1964; Rabin and Thaler 2001). The demand for insurance arises because a dollar of premium paid in the no-loss state when wealth is high is worth less (in utility) than a dollar of *payment* received in the loss-state (when wealth is low). Even if one's flat screen TV set blows up one year after purchase, the loss of one or two thousand dollars should not "move" the marginal utility of wealth by very much for most people. Demand for EWs thus cannot plausibly be attributed to classical risk aversion: whatever the reason for buying an EW, nobody could be risk-averse enough to justify the purchase for a small-value item. Indeed, recent experimental evidence (Huysentruyt and Reed 2010; Jindal 2013) suggests that EW purchases are much more driven by loss- or regret-aversion, rather than a rational calculation of the expected utility gained from the purchase.

The second reason why EWs are such a bad deal is that they are exceedingly expensive. Estimates vary, but profits on EWs are many times higher than on standard insurance policies, and loss ratios (payouts to premiums collected) are often staggeringly low (Baker and Siegelman 2013; UK Competition Commission 2005).



There are thus two puzzles about EWs: why do people want them at all, and even if they do, why don't the forces of competition push the profits on EWs down to normal levels. Behavioral economics offers compelling answers to both these questions.

#### 4.1.1 *What Explains the Demand for EWs?*

Behavioral analysis offers a wide range of explanations for consumers' irrational purchase of extended warranties. Indeed, there are so many plausible stories that distinguishing between them is quite difficult. Fortunately, however, the normative conclusions appear to be independent of the particular mechanism involved.

One possible story involves "regret aversion," under which people (know now that they) will feel bad in the future if a choice they have made today works out badly, even if it was appropriate at the time it was made. Theoretical work by Michael Braun and Alexander Muermann (2004) shows that regret aversion leads people to purchase insurance for low-value losses: when insurance is available, is not purchased, and a loss occurs, a regret-averse person will feel this loss very heavily, since she could have chosen to buy insurance but did not.

Another variation on this theme is the idea that utility is not simply a function of final wealth, but depends on whether that final wealth is framed or experienced as a gain or a loss relative to some reference point (Kahneman and Tversky 1979). According to Kahneman and Tversky's prospect theory, this reference dependence is accompanied by loss aversion and diminishing sensitivity. Loss aversion is the phenomenon that losses loom larger than gains: people hate to lose more than they like to gain (chapter 11 by Zamir in this volume). Parting with goods that are held for exchange, such as money, is not, however, perceived as a loss, but rather as a "cost" (Novemsky and Kahneman 2005). Diminishing sensitivity means that people value the first dollar of a gain the most and each additional dollar of gain less. At the same time, people hate the first dollar of a loss more than any additional dollar. In other words, they have a declining marginal *disutility* of loss that mirrors their declining marginal utility of gains. All this means that people will often pay dearly to avoid even a small loss (Johnson et al. 1993). In the add-on insurance context, that translates into paying what feels like a small additional cost to avoid the emotional distress associated with a larger future loss.

There are still other behavioral explanations for the irrational willingness to buy extended warranties, including those based on nonlinear "probability weighting" (Kahneman and Tversky 1979; Barberis 2013) (meaning that small loss probabilities are exaggerated relative to their true value, while larger ones are understated); and of course several explanations could be operating at the same time. Distinguishing between them is extraordinarily difficult, although one recent attempt to do so using experimental data and statistical methods concludes that loss aversion is the most significant driver of EW purchases (Jindal 2013). Our sense is that while a precise understanding of consumers' motivations is a worthwhile objective, what really matters, at least in this context, is that EWs are a bad deal for a rational consumer.

#### 4.1.2 *What Explains the High Profit Margins on EWs?*

For an explanation of why competition among retailers fails to reduce the profit margin on EWs, we rely on the well-known “shrouded pricing” model of two-stage purchases first developed by Gabaix and Laibson (2006). We summarize that model here, stressing its prediction that when some actors are subject to a plausible behavioral anomaly, inefficient and discriminatory terms *can* survive in equilibrium. It is important to note that this result is at odds with the traditional story of equilibrium in markets with rational but imperfectly informed consumers (Schwartz and Wilde 1979, 653). In the standard account, “the presence of at least some consumer search in a market creates the possibility of a ‘pecuniary externality’: persons who search sometimes protect nonsearchers from overreaching firms.” Moreover, in Schwartz and Wilde’s simulation model, if at least one-third of consumers undertake comparison shopping, the market price will be close to the competitive price in market where all consumers are informed.

In the shrouding model, a consumer has to make an initial purchase, and then *optionally* makes a secondary purchase that is somehow tied to the first. Gabaix and Laibson use examples such as a laser printer and replacement cartridges, a hotel room and telephone charges, or a car rental and a collision damage waiver. There are two kinds of consumers in their model—“myopes,” who don’t think about the possibility of future “add-ons” when they make their initial purchase, and “sophisticates,” who do. The initial purchase is made in a competitive market, where all stage 1 prices of all sellers are completely observable; but the first purchase exposes the buyer to a subsequent purchase from the same seller, in a potentially noncompetitive market in which the price is unobservable at the time the initial purchase is made (unless one inquires about it).

As Gabaix and Laibson observe, the second-stage price is often significantly above the marginal cost of providing the good or service. That is certainly the case for extended warranties, where sellers earn margins that are unheard of in virtually any other line of insurance. We think it is helpful to think of the second-stage purchase as taking place in a “situational monopoly” in which the seller has a captive market for that part of the purchase. One could presumably buy an extended warranty separately from the primary purchase, but this turns out to be very rare in practice, with the result that extended warranties are sold at decidedly supracompetitive, monopoly-like prices. The shrouded pricing model provides an explanation for why.

Suppose a firm tries to compete by offering a lower second-stage price than its rival—for example, on extended warranties—and by alerting potential customers to the fact that its rivals charge more (“come buy from us—we charge less for our warranties”). Doing so has several consequences. First, the fact that the overall market is competitive means that the firm offering cheaper warranties would have to charge a higher price for the first-stage product—otherwise, the discounter would earn negative profits and would prefer to exit the industry. Second, the discounter’s announcement educates its rivals’ sophisticated consumers, alerting them to the fact that cheaper warranties are possible. But ironically, this means that rivals’ customers will *all* prefer to

stay where they are, rather than switch to the discounter. Sophisticated customers will want to stick with the rival to obtain the lower base charge, and will avoid the rival's high add-on charges by substituting a competitively supplied extended warranty for that offered by the seller; or, better yet, by not buying one at all and relying instead on savings or a credit card (which provides the liquidity needed to purchase a replacement and, in some cases, may include limited warranty protection on purchases made with the card) to replace the product if it breaks. Importantly, however, this advertising will have no effect on rivals' myopic consumers, who aren't paying attention to the second-stage transaction at all. Thus, competitive attempts to unmask rivals' high add-on prices will only succeed in transferring benefits from the rival to the rival's sophisticated customers, and will not do anything for the firm providing the educational information at all. Hence, there will be no reason for any firm to unmask its rivals' high add-on fees, which can then persist in equilibrium.

#### 4.1.3 *Insights from the Shrouded Pricing Model*

The shrouding model offers several important insights for the application of behavioral economics to the regulation of insurance. Most significantly, it shows how behavioral "flaws" don't just influence the consumer's decision about what/how much to buy. These flaws also shape the structure of competition between firms and the resultant market equilibrium. An analysis that focuses only on consumers' deviations from perfect rationality (or nonstandard preferences) will miss the important properties of the equilibrium that results. Sadly, there is thus no shortcut from behavioral anomaly directly to policy recommendations: rather, as the previous examples also demonstrate, the behavioral anomalies have to be inserted into an overall model of market functioning to predict how policy can influence welfare.

The shrouding model also helps explain why the enhanced disclosure approach to extended warranty overcharges proposed by the UK Competition Commission failed so dramatically. After an impressive empirical analysis of the market for EWs in the UK, the Competition Commission decided to require advertising of the extended warranty price along with the price of the covered product, thereby allowing consumers to shop on the basis of the combined price; the Commission also proposed further reforms of the sales process designed to reduce the likelihood that customer would be pressured into buying an EW (UK Competition Commission 2003). Along with some other reforms, the Commission's proposals were adopted by regulation, effective April 2005.<sup>4</sup>

Yet profits from extended warranties on consumer electronic products in the UK continue to be very high, despite the reforms, and the UK Office of Fair Trading still sees the market as "unfair and uncompetitive" (Neate 2011). Just as the shrouded pricing model would predict, disclosure did not work. True, prices of extended warranties

<sup>4</sup> <http://www.competition-commission.org.uk/inquiries/completed/2003/warranty/index.htm>; <http://www.legislation.gov.uk/uksi/2005/37/contents/made> (the regulation as adopted allowed for a 45-day cancellation period).

have declined at traditional retailers since the reforms. But that appears to be the result of competition from internet retailers and big box stores (Office of Fair Trading 2008). The Office of Fair Trading's follow-up investigation concluded that disclosure was not working and recommended, instead, an information technology solution that would eliminate the situational monopoly (Office of Fair Trading 2012). British retailers recently accepted that recommendation as an "agreed remedy," perhaps to avoid the ban that we recommend for extended warranties in the add-on context.

Exactly why disclosure failed is ultimately an empirical question, but the shrouding model would say that myopic consumers ignored the disclosures, while rational consumers reasoned that lower warranty prices must mean higher up-front prices, so that disclosure would not in fact enhance competition over customers.

Equilibrium analysis also bears on the paternalism problem, voiced by Dan Schwarcz (2010). Suppose we concede that consumers are not "mistaken" in many insurance purchasing decisions and that, instead, they are motivated to purchase credit life insurance, flight insurance, collision damage waivers, or extended warranties by genuine (albeit "nonstandard") fears or anxieties. It does not follow that consumers should overpay for the insurance they purchase, as the shrouding model predicts and the evidence strongly suggests is the case. In other words, an equilibrium behavioral analysis might still suggest a market failure that regulation could potentially address, even if insurance is purchased for "legitimate but non-standard" reasons such as regret- or loss-aversion. The market failure arises not from consumer motivations per se; it arises from the way such motivations shape the resultant market equilibrium and reduce the ability of competitive market forces to present consumers with prices that closely track the cost of providing the service (see also Baker and Siegelman 2013).

## 4.2 Low-Deductible Homeowners Insurance

Consumers' choice of deductible in homeowner's insurance provides a second compelling example of the kinds of problems we have been discussing. Research by Justin Sydnor (2010) conclusively demonstrates that many policyholders choose deductibles that are *much* too low to be justified as the decision of a rational, risk-averse actor. Sydnor uses data from one large insurer to demonstrate that 83% of consumers choose a deductible that is dramatically too small to be justified by any reasonable level of risk aversion or future expected claims. For example, many consumers chose a \$500 deductible, rather than the \$1,000 deductible they might have picked instead. Given typical claiming rates, the average expected monetary benefit from the additional coverage was only about \$20, but its additional cost was about five times more than that. In other words, consumers paid \$100 to receive an expected \$20 monetary benefit (Sydnor 2010, 196).

Sydnor's was an observational study, which limited his ability to explain precisely *why* consumers were willing to overpay for a low deductible. But he suggests this finding might be explained by a number of behavioral anomalies, but an obvious candidate

would be loss aversion: paying a higher premium in return for a lower deductible “just doesn’t feel” like a loss, in the same way that paying out of pocket for an uncovered loss (that would have been covered had the policyholder chosen a lower deductible) does (see also Johnson et al. 1993).

To justify the lower deductible, a rational consumer would have to have a utility function that was so astronomically risk-averse that he would almost literally never be able to get out of bed. In quantitative terms, buying the lower deductible is a rational economic decision only if one’s coefficient of relative risk aversion is between 1,840 and 5,064. Yet empirical studies estimate plausible values for the coefficient of relative risk aversion to be in the single-digit range. Someone with a coefficient of relative risk aversion of 5,000 would turn down a bet that offered a 50/50 chance of either losing \$1,000 or gaining *any amount of money* (including, say \$1,000,000,000,000). (Sydnor 2010, table 3, 190). Instead, Sydnor’s preferred explanation for the purchase of unreasonably low deductibles is that consumers have inconsistent and imperfectly rational preferences that do not match those in the standard economic account. “Feelings about money given up for a purchase are segregated from attitudes towards surprise losses,” such that “loss aversion affects attitudes towards money paid when an accident happens (i.e., the deductible) but not the amount of money paid up front for the policy” (Sydnor 2010, 196).

Whatever the explanation for this choice, the purchase of “excess” deductibles appears to be costly to consumers: Sydnor estimates that other things equal, “homeowners could expect to save roughly \$4.8 billion per year by holding the highest available deductible.” (One might think of the \$4.8 billion as money well spent *given consumers’ actual, if inconsistent, preferences*. Alternatively, the \$4.8 billion might be characterized as a cost of irrationality that society ought to take steps to overcome.). But as Sydnor points out, this analysis can be seriously misleading as a guide for regulation, because it ignores the way markets equilibrate. Consistent with competition among suppliers, the insurer he studied did not appear to earn excess profits on its low-deductible policies, even though consumers “overpaid” for these policies relative to the expected value of the low deductible. This is because low-deductible consumers had higher claim rates, presumably due to the presence of adverse selection. The low-deductible consumers, who had private information about their own elevated likelihood of making a claim, chose policies that reflected this information, even though the additional expected claims were not “worth” the cost of the additional coverage. In fact, those with a \$500 deductible had about a 50% higher claim rate (between 3% and 3.5% per year) than those with a \$1,000 deductible (only about 2% of whom made a claim each year), by various measures that controlled for the fact that people with a \$1,000 deductible cannot make a claim for a \$900 loss (Sydnor 2010, 198; it is important to control for the fact that those with a lower deductible can make claims for amounts between \$500 and \$1,000 that those with a higher deductible cannot; thus, it is appropriate to use the rate of claims in excess of the higher deductible for this comparison). Thus, if the low-deductible policies were to be eliminated, the equilibrium would look very different and might not exist at all.

I may be able to get a better view at the ball game if I stand up, but this does not imply that *everyone* can simultaneously get a better view if we all do so. Similarly, Sydnor (2010, 198) concludes that “[i]ndividual consumers could benefit financially by avoiding over-insuring modest risk. However, if all homeowners changed their behavior, the company would likely need to raise insurance costs or create a new higher deductible in order to separate the more and less risky consumers. . . . If all consumers had standard risk preferences, the new market equilibrium would not necessarily be welfare-improving for the customers.”

## 5 CONCLUSION

---

As we noted at the outset, extended warranties and low-deductible homeowners insurance are ideal examples for evaluating the potential contributions of behavioral economics to insurance regulation. The market outcomes sharply diverge from the predictions of expected utility theory, and we understand the institutional context sufficiently well to conclude that the divergence reflects nonrational behavior. In addition, behavioral economics provides a reasonably well-grounded set of explanations for the observed behavior, posing a clear challenge to expected utility as a positive theory. Both examples thus present difficult, practical, and generalizable problems for regulators: how should policymakers respond when consumers apparently “want” (or at least, are willing to pay for) something that a rational person would not choose to buy? What, if any, forms of regulation are likely to be effective and desirable under these conditions?

We conclude that even though it does not explain actual consumer behavior very well, expected utility theory may serve as an acceptable descriptive basis for *normative* policy evaluation. Insurance consumers are clearly willing to pay much too much for products they shouldn’t want at all, and behavioral economics offers powerful explanations for why this should be so. As has long been noted (Kunreuther and Pauly 2005; Kunreuther, Pauly, and McMorrow 2013; Cutler and Zeckhauser 2004) there are parallel anomalies involving the opposite form of behavior: consumers fail to purchase certain kinds of insurance for large losses, even when that insurance is available at subsidized rates that make it better than actuarially fair. We do not discuss these anomalies here, but we believe our approach is applicable in those contexts as well.

How, then, should regulators respond to such irrational behavior? One possibility is simply to ignore consumers’ motivations, relying on a “revealed preference” approach that presumes that if consumers are willing to pay for something, their welfare is enhanced by purchasing it, even if no rational person would make this choice. We think that insurance products that so dramatically fail the rational expected utility maximizer’s cost-benefit calculus are not worthy of this respect. They are bad deals, and we suspect that the high-pressure and deceptive tactics used to sell them explain a



large fraction of the demand for such products (on the exploitation of consumer biases by suppliers; see generally chapter 18 by Bar-Gill in this volume).

For extended warranties we advocate paternalist regulation of the strongest kind: prohibiting their sale, at the very least in the contexts in which people presently are most likely to buy them (Baker and Siegelman 2013). (For the full explanation, readers will have to consult our paper. One intuition is that a mandatory rule eliminates regret aversion because there is nothing to regret any more.) For low-deductible homeowners' insurance, we advocate leaving the market alone. The difference is not because people are making good, or even well-informed, choices in one context but not the other. The difference lies in the equilibrium effects of the choices in these two markets.

As these examples illustrate, the most significant regulatory payoffs from the encounter between behavioral economics and insurance have not come from a more precise understanding of the motives for buying insurance or from advances in identifying what constitutes "good" insurance or the ultimate goals of insurance regulation.<sup>5</sup> Rather, the benefits have come from advances in the understanding of the equilibrium that results when real—incompletely rational—people buy insurance.

A recent paper by Handel (2013) beautifully illustrates the central theme of this chapter: deviations from rational behavior that might be welfare-reducing in standard market settings (or when considered in isolation) may actually be welfare-enhancing in insurance market equilibrium. Handel examines "inertia" in individuals' choice of employer-provided health insurance plans. (Although his analysis does not identify a particular cognitive failure that gives rise to this inertia, it is not difficult to imagine several behavioral explanations.) He finds compelling empirical evidence that people tend to stick with a given plan, even if other offerings are clearly better for them: such inertia "causes an average employee to forgo \$2,032 annually," a substantial fraction of the \$4,500 that an average employee's family spends each year (Handel 2013, 2645). Inertia obviously leads to individual welfare losses. But its upside is that it reduces adverse selection, precisely because it retards consumers' tendency to utilize their informational advantage in choosing the insurance plan that is best for them. A welfare analysis that combines both consumer inertia and adverse selection requires a model of how insurers would alter pricing in response to selection pressures. After developing and tested such a model, Handel concludes that "where insurance prices endogenously respond to different enrollment and cost patterns, . . . [an intervention that] reduces inertia by three-quarters. . . improves consumer choices conditional on prices, but. . . also exacerbates adverse selection, leading to a 7.7% *reduction* in welfare" (Handel 2013, 2646; *emph. added*). The bottom line is simple, if paradoxical: given the complexities of insurance market equilibria, behavioral failings can actually increase welfare. There is

<sup>5</sup> We do not mean to imply by this that behavioral research is, or should be, directed at determining what is "good" insurance. Behavioral decision research often lacks such a normative ambition.



thus no warrant for believing that “correcting” irrational behavior in insurance markets is justified on efficiency grounds.

Others have made this point before, but it is worth emphasizing: behavioral decision research has normative implications, but those implications are largely in the realm of means not ends. Eventually, behavioral economics may lead to a new understanding of what constitutes “good” insurance, but the “behavioral” part of behavioral economics has not produced a widely accepted alternative to expected utility theory as a normative guide. (This is not a criticism of behavioral economics.) For the moment at least, what the behavioral turn can do is help policymakers design better tools for achieving the ends that any pertinent normative theory identifies. For the insurance field, the payoff lies in devising ways to help consumers choose good insurance products, identifying situations in which they are so unlikely to make good choices that stronger regulation is justified, and, with appropriate attention to equilibrium analysis, guiding policymakers in designing regulatory strategies, such as those addressing the extended warranty market.

What we have learned from working through these examples extends beyond insurance. Behavioral economics does provide scope for welfare-enhancing interventions. But once we admit the existence of nonstandard motivations or imperfect reasoning, those interventions must be very carefully tailored to particular circumstances in order to be effective. There is no simple route from identifying a nonstandard motivation or a “flaw” in reasoning to recommending an appropriate regulatory policy. Put perhaps too simply, we need psychology to identify how people reason, economics to understand the consequences of that behavior for market equilibrium, and law and other disciplines that reward detailed institutional knowledge to incorporate these insights into regulatory strategies that have a chance of moving the market toward a new, welfare-enhancing equilibrium.

## REFERENCES

- Abraham, Kenneth. 1986. *Distributing Risk: Insurance, Legal Theory, and Public Policy*. New Haven: Yale University Press.
- Affordable Care Act, § 1501(a)(2)(I), as amended by § 10106.
- Akerloff, George A. 1970. The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics* 84:488–500.
- Alborn, Timothy. 2009. *Regulated Lives: Life Insurance and British Society 1800–1914*. Toronto: University of Toronto Press.
- ALI Liability Insurance Project Preliminary Draft No. 1, 2013.
- Arrow, Kenneth J. 1963. Uncertainty and the Welfare Economics of Medical Care. *American Economic Review* 53:941–73.
- Arrow, Kenneth J. 1971a. Insurance, Risk and Resource Allocation. In *Essays in the Theory of Risk Bearing*, 134–79. Chicago: Markham.
- Arrow, Kenneth J. 1971b. The Theory of Risk Aversion. In *Essays in the Theory of Risk Bearing*, 90–109. Chicago: Markham.

- Baker, Tom. 1996. On the Genealogy of Moral Hazard. *Texas Law Review* 75:237–92.
- Baker, Tom. 2003. Containing the Promise of Insurance: Adverse Selection and Risk Classification. In *Risk and Morality*, ed. R. V. Ericson and A. Doyle, 258–83. Chicago: University of Chicago Press.
- Baker, Tom, and Peter Siegelman. 2010. Tontines for the Invincibles. *Wisconsin Law Review* 2010:79–120.
- Baker, Tom and Peter Siegelman. 2013. You Want Insurance With That? Using Behavioral Economics to Protect Consumers from Add-On Insurance Products. *Connecticut Insurance Law Journal* 20:1–60.
- Barberis, Nicholas C. 2013. Thirty Years of Prospect Theory in Economics: A Review and Assessment. *Journal of Economic Perspectives* 27:173–96.
- Beal, Eric D. 2001. Posner and Moral Hazard. *Connecticut Insurance Law Journal* 7:81–102.
- Ben-Shahar, Omri and Carl E. Schneider. 2011. The Failure of Mandated Disclosure. *University of Pennsylvania Law Review* 159:647–749.
- Breach, Anat, and Donald J. Brown. 2010. Affective Decision Making: A Theory of Optimism Bias. Federal Reserve Bank of Boston Working Paper 10-16. Available at <http://www.bos.frb.org/economic/wp/index.htm>.
- Braun, Michael, Peter S. Fader, Eric T. Bradlow, and Howard C. Kunreuther. 2006. Modeling the “Pseudodeductible” in Insurance Claims Decisions. *Management Science* 52:1258–72.
- Braun, Michael, and Alexander Muermann. 2004. The Impact of Regret on the Demand for Insurance. *Journal of Risk and Insurance* 71:737–67.
- Bruce, J., and J. E. Johnson. 1996. Decision Making under Risk: Effect of Complexity on Performance. *Psychological Report* 79:67–76.
- Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O’Donoghue, and Matthew Rabin. 2003. Regulation for Conservatives: Behavioral Economics and the Case for Asymmetric Paternalism. *University of Pennsylvania Law Review* 151:1211–54.
- Camerer, Colin F. and Howard C. Kunreuther. 1989. Decision Processes for Low Probability Events: Policy Implications. *Journal of Policy Analysis and Management* 8:565–92.
- Chan, Cherie. 2012. *Marketing Death: Culture and the Making of a Life Insurance Market in China*. New York: Oxford University Press.
- Chandler, Seth. N.D. Insurance Regulation. In *Encyclopedia of Law and Economics*. Available at <http://encyclo.findlaw.com/5700book.pdf>.
- Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2011. Who Is (More) Rational? National Bureau of Economic Research Working Paper no. 16971.
- Clark, Geoffrey. 1999. *Betting on Lives: The Culture of Life Insurance in England 1695–1775*. Manchester: Manchester University Press.
- Cohen, Alma, and Peter Siegelman. 2010. Testing for Adverse Selection in Insurance Markets. *Journal of Risk and Insurance* 77:39–84.
- Cutler, David M., and Richard J. Zeckhauser. 1998. Adverse Selection in Health Insurance. In *Frontiers in Health Policy Research*, vol. 1, ed. Alan M. Garber, 77–116. Cambridge, MA: MIT Press.
- Cutler, David M., and Richard A. Zeckhauser. 2004. Extending the Theory to Meet the Practice of Insurance. In *Brookings-Wharton Papers on Financial Services*, ed. Robert Litan and Richard Herring. Available at <http://fic.wharton.upenn.edu/fic/brookings.htm>.
- Daston, Lorraine. 1988. *Classical Probability in the Enlightenment*. Princeton, NJ: Princeton University Press.

- Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. Are Risk Aversion and Impatience Related to Cognitive Ability? *American Economic Review* 100:1238–60.
- Feldstein, Martin S. 1973. The Welfare Loss of Excess Health Insurance. *Journal of Political Economy* 81:251–80.
- Fredrick, Shane. 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives* 19:25–40.
- Gabaix, Xavier, and David Laibson. 2006. Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets. *Quarterly Journal of Economics* 121:505–40.
- Handel, Benjamin R. 2013. Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. *American Economic Review* 103:2643–82.
- Harel, Alon, and Yuval Procaccia. 2009. On the Optimal Regulation of Unread Contracts. Available at SSRN: <http://ssrn.com/abstract=1473770>.
- Heimer, Carol. 1985. *Reactive Risk and Rational Action*. Berkeley: University of California Press.
- Hendel, Igal, and Alessandro Lizzeri. 2003. The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance. *Quarterly Journal of Economics* 118:299–327.
- Hogarth, Robin M., and Howard C. Kunreuther. 1995. Decision Making under Ignorance: Arguing with Yourself. *Journal of Risk and Uncertainty* 10:15–36.
- Hsee, Christopher K., and Howard C. Kunreuther. 2000. The Affection Effect in Insurance Decisions. *Journal of Risk and Uncertainty* 20:141–60.
- Huysentruyt, Marieke, and Daniel Read. 2010. How Do People Value Extended Warranties? Evidence from Two Field Surveys. *Risk and Uncertainty* 40:197–218.
- Jackson, Howell. 1999. Regulation in a Multi-sectored Financial Services Industry. *Washington University Law Quarterly* 77:319–97.
- Jaffee, Dwight. 2006. Monoline Restrictions, with Applications to Mortgage Insurance and Title Insurance. *Review of Industrial Organization* 26:83–108.
- Jindal, Pranav. 2013. Risk Preferences and Demand Drivers of Extended Warranties. Available at SSRN: <http://ssrn.com/abstract=2196033>.
- Johnson, Eric, John Hershey, Jacqueline Meszaros, and Howard C. Kunreuther. 1993. Framing, Probability Distortions and Insurance Decisions. *Journal of Risk and Uncertainty* 7:35–51.
- Kahneman, Daniel, and Amos N. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47:263–92.
- Karla, Ajay, Baohong Sun, and Tao Chen. 2009. Why Do Consumers Buy Extended Service Contracts? *Journal of Consumer Research* 36:611–23.
- Keller, Carmen, Michael Siegrist, and Heinz Gutscher. 2006. The Role of the Affect and Availability Heuristics in Risk Communication. *Risk Analysis* 26:631–39.
- Krantz, David H., and Howard C. Kunreuther. 2007. Goals and Plans in Decision Making. *Judgment and Decision Making* 2:137–68.
- Kunreuther, Howard C., Mark V. Pauly, and Stacey McMorro. 2013. *Insurance and Behavioral Economics: Improving Decisions in the Most Misunderstood Industry*. New York: Cambridge University Press.
- Kunreuther, Howard C., Warren Sanderson, and Rudolf Vetschera. 1985. A Behavioral Model of the Adoption of Protective Activities. *Journal of Economic Behavior and Organization* 6:1–15.
- Langer, Jesse A. 2001. Deciphering Posner's Insurance Jurisprudence. *Connecticut Insurance Law Journal* 7:1–6.

- Loomes, Graham, and Robert Sugden. 1982. Regret Theory: An Alternative Theory of Rational Choice under Uncertainty. *Economic Journal* 92:805–24.
- National Association of Insurance Commissioners, Exchanges (B) Subgroup. 2011. Adverse Selection Issues and Health Insurance Exchanges Under the Affordable Care Act. Available at [http://www.naic.org/documents/committees\\_b\\_110622\\_adverse\\_selection.pdf](http://www.naic.org/documents/committees_b_110622_adverse_selection.pdf).
- Neate, Rupert. 2011. OFT to Look into Extended Warranties. *Daily Telegraph*, April 15, Business Section, 3.
- Novemsky, Nathan, and Daniel Kahneman. 2005. The Boundaries of Loss Aversion. *Journal of Marketing Research* 42:119–28.
- Office of Fair Trading. 2008. Evaluating the Impact of the Supply of Extended Warranties on Domestic Electrical Goods Order 2005. Available at [http://www.offt.gov.uk/shared\\_offt/reports/Evaluating-OFTs-work/oft1024.pdf](http://www.offt.gov.uk/shared_offt/reports/Evaluating-OFTs-work/oft1024.pdf).
- Office of Fair Trading. 2012. Extended Warranties on Domestic Electrical Goods, OFT 1403. Available at [http://www.offt.gov.uk/shared\\_offt/markets-work/OFT1403.pdf](http://www.offt.gov.uk/shared_offt/markets-work/OFT1403.pdf).
- Pratt, John W. 1964. Risk Aversion in the Small and in the Large. *Econometrica* 32:122–136.
- Priest, George L. 1981. A Theory of the Consumer Warranty. *Yale Law Journal* 90:1297–352.
- Rabin, Michael, and Richard H. Thaler. 2001. Anomalies: Risk Aversion. *Journal of Economic Perspectives* 15:219–32.
- Rothschild, Michael, and Joseph Stiglitz. 1976. Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *Quarterly Journal of Economics* 90:629–49.
- Sandroni, Alvaro, and Francesco Squintani. 2007. Overconfidence, Insurance, and Paternalism. *American Economic Review* 97(5): 1994–2004.
- Schwarcz, Daniel. 2010. Regulating Consumer Demand in Insurance Markets. *Erasmus Law Review* 3:23–45.
- Schwarcz, Daniel. 2011. Re-evaluating Standardized Insurance Policies. *University of Chicago Law Review* 78:1263–349.
- Schwartz, Alan, and Louis L. Wilde. 1979. Intervening in Markets on the Basis of Imperfect Information: A Legal and Economic Analysis. *University of Pennsylvania Law Review* 127:630–35.
- Segal, David. 2012. *Selling It with Extras, or Not at All*. *New York Times*, September 9, B4.
- Shavell, Steven. 1979. On Moral Hazard and Insurance. *Quarterly Journal of Economics* 92:541–62.
- Shoemaker, Paul J. H., and Howard C. Kunreuther. 1979. An Experimental Study of Insurance Decisions. *Journal of Risk and Insurance* 46:603–18.
- Slovic, Paul, Baruch Fischhoff, Sarah Lichtenstein, B. Corrigan, and B. Combs. 1977. Preference for Insuring against Probable Small Losses: Insurance Implications. *Journal of Risk and Insurance* 44:237–58.
- Spence, A. Michael. 1977. Consumer Misperceptions, Product Failure and Producer Liability. *Review of Economic Studies* 44:561–72.
- Sulzle, Kai, and Achim Wambach. 2005. Insurance in a Market for Credence Goods. *Journal of Risk and Insurance*, 72:159–76.
- Sydnor, Justin. 2010. (Over)insuring Modest Risks. *American Economic Journal: Applied Economics* 2:177–99.
- Tversky, Amos N., and Daniel Kahneman. 1982. Availability: A Heuristic for Judging Frequency and Probability. In *Judgment under Uncertainty: Heuristics and Biases*, ed.

Daniel Kahneman, Paul Slovic, and Amos N. Tversky, 163–89. Cambridge: Cambridge University Press.

UK Competition Commission. 2003. A Report on the Supply of Extended Warranties on Domestic (Household) Electrical Goods within the UK. Vol. 1.

Willis, Lauren E. 2008. Against Financial-Literacy Education. *Iowa Law Review* 94:197–285.

## CHAPTER 20

---

# THE END OF CONTRACTARIANISM?

## *Behavioral Economics and the Law of Corporations*

---

KENT GREENFIELD

### 1 INTRODUCTION

---

FOR a generation, the law of corporations depended on, and sprang from, a notion of economic rationality. This rationality took as its touchstone the efficiency of the marketplace (especially the securities market) and the predictability of the utility-maximizing behavior of the various actors. These assumptions undergirded both the doctrinal innovations and the academic conceptualizations of the field (Easterbrook and Fischel 1991), leading to a prioritization of contractarian reasoning in chancery courts and corporate law classrooms (Marnet 2008). The judicial enforcement of traditional fiduciary duties melted away, as markets themselves came to be seen as the primary method of keeping management careful and loyal. Disclosure was vested with the mantle of favored regulatory mechanism, since the dominant assumption was that investors armed with information needed no other protection from corporate negligence or malfeasance.

Behavioral economics began to be taken seriously in the legal academy in the last decade of the twentieth century (Jolls, Sunstein, and Thaler 1998; Langevoort 1998; Korobkin and Ulen 2000), and by the early 2000s was beginning to gain traction in corporate law scholarship (Langevoort 2000; Blair and Stout 2001; Greenfield 2002). Though corporate and securities law was perhaps the last bastion in the legal academy of the assumptions of neoclassical economics (Arlen, Spitzer, and Talley 2002), it is safe to say that the global financial crisis of 2007–8 finally marked the end of the glory days of homo economicus. If a stalwart of the rationality school such as former chairman of the US Federal Reserve Alan Greenspan was forced to admit that he had “found a flaw” in his theory of the free market (Andrews 2008), then few absolutists were left indeed. Now, behavioral economics can claim victory, at least in the academy. As Russell Korobkin opined in 2011, “the battle to separate the economic analysis of legal rules

and institutions from the straightjacket of strict rational-choice assumptions has been won” (Korobkin 2011, p. 1655). The evidence of the persistence and ubiquity of so-called irrationalities is so pervasive, replicable, and defensible that the primary debate is no longer between behaviorists and those who would defend “the traditional faith in individual optimization as a core analytical assumption of legal analysis” (Korobkin 2011, p. 1656). Instead the debate is over what to do with, and about, the spoils of victory.

This chapter will offer some thinking in that regard in connection with corporate and securities law. To begin, the chapter will describe some of the ways in which economic rationality influenced corporate law doctrine and scholarship during its heyday. This backward nod is helpful in defining possible implications for the weakening of the rationality assumptions later. Section 2 focuses on the difficulty of developing behavioral research in the area of corporate governance, arguing that the main behavioral innovation is a move away from libertarianism toward regulatory “agnosticism.” In section 3, I will discuss what I believe to be the most profound potential implication of behavioral scholarship on corporate governance: using insights on the decision-making of groups to bolster the argument that board homogeneity is a danger, and that increased board diversity of various kinds is likely to improve board performance.

## 2 CHOICE AND VOLUNTARINESS IN CONTRACTARIAN CORPORATE LAW

---

Behavioral economics presented a profound challenge to corporate law, since much of it was based on “the Coasean prediction” that the law can and should presumptively rely on the decisions of private parties “to allocate rights and obligations optimally” (Arlen, Spitzer, and Talley 2002, p. 2). This contractarian conception of corporate law was dominant for several decades, and while there were scholars who offered dissents (Brudney 1997) or caveats (Bebchuk, Cohen, and Ferrell 2002), the trends and fundamental ideas were clear.

The corporation was seen as a nexus of contracts among a host of different contributors. Shareholders were not seen as owners to which fiduciary duties were owed because of their status; instead they were simply contributors of capital who, in a sense, gained legal protection through negotiation and purchase. Managerial obligations were derived from explicit, implicit, or default contractual terms between management and shareholders, not from any conception of fiduciary obligation deriving from property, agency, or trust law concepts. Law, it was argued, should not dictate the obligations among a corporate contract’s parties because each party was assumed to know her own interests and to protect them through bargaining and exchange. Developments in corporate law and in corporate charters were assumed to trend toward efficiency, since any inefficiency would cause participants in those arrangements to change the terms of the



bargain over time in order to maximize their utility and avoid losses (Easterbrook and Fischel 1991). The dedication to private ordering was sufficiently deep that corporate law came to be even more contractarian than contract law itself (Brudney 1997), in that it recognized fewer behavior-based exceptions to private ordering such as unconscionability (Greenfield 2011b).

## 2.1 Judicial Deference

One of the main implications of contractarian theory was the pervasive deference courts showed to the decisions of corporate executives. Courts stepped aside, convinced that the arrangements among the various parties were voluntary and that the market itself would correct any mistakes. Perhaps the best statement of these judicial assumptions appears in Judge Ralph Winter's famous opinion for the Second Circuit in *Joy v. North* (692 F.2d. 880 (1982)), one of the earliest case-law articulations of the contractarian model (an opinion still appearing in many leading corporate law case-books). Explaining why shareholders should not be able to win a fiduciary-duty suit brought against negligent directors, Judge Winter posited that shareholders had chosen to take on the risk of directors' negligent behavior: "[S]hareholders to a very real degree voluntarily undertake the risk of bad business judgment" (id., p. 885). They had expressed their choice simply by buying the stock. As Judge Winter stated (id.):

Investors need not buy stock, for investment markets offer an array of opportunities less vulnerable to mistakes in judgment by corporate officers. Nor need investors buy stock in particular corporations. In the exercise of what is genuinely a free choice, the quality of a firm's management is often decisive and information is available from professional advisors. Since shareholders can and do select among investments partly on the basis of management, the business judgment rule merely recognizes a certain voluntariness in undertaking the risk of bad business decisions.

Notice the powerful contractarian view imbedded in this argument. The shareholders had not entered into an agreement waiving their right to sue the directors for breach of fiduciary duty. The question of what the shareholders had agreed to—that is, the level of fiduciary duty contained in the corporate "contract" between shareholders and management—was the question to be answered by the case. Judge Winter answered it by simply noting that the shareholders had purchased stock. That is, investors entered into a contract because they acted in a way that brought about consequences they might have anticipated. In other words, because people know that some managers make mistakes and because people are rational actors, we can assume people accept the risk of managerial mistakes when they purchase stock in a company. Shareholders would receive a windfall if they were allowed to recover for those mistakes.

This opinion embodied the dominant legal-economic assumptions of the time. Three decades on, we can now easily recognize the simplistic and problematic view

of human rationality and voluntariness on which this line of argument depends: we know what people want because their actions reflect what they want. This view of revealed-preferences-as-rationality is so thin that Judge Posner could say in early editions of his seminal book, “[I]t would not be a solecism to speak of a rational frog” (Posner 2003, p. 17).

In a way, the thinness of this view of rationality was its power—courts would not need to probe beyond the straightforward query of whether the corporation coerced or fraudulently induced an individual into behaving a certain way. If not, and barring externalities, market asymmetries, or other garden-variety market failures, then courts should let things lie.

Given this thin view of rationality and voluntariness, the fact that shareholders invested in the firm was evidence that they assumed the risk of whatever managerial malfeasance occurred. If they thought malfeasance likely, then they should either not have invested or they should have contracted for protections. If they did not protect themselves in such ways, they should not expect courts to rescue them after something went wrong.

## 2.2 Academic Defenses

It was not only courts that were taken with the efficiency of the market and the reliability of revealed preferences. These concepts also undergirded the leading academic defenses of the contractarian model of corporate law. In the words of Michael Jensen and William Meckling, “Contractual relations are the essence of the firm” (Jensen and Meckling 1976, p. 310). University of Chicago professor Daniel Fischel argued in one of his first descriptions of the nexus-of-contracts version of corporate law that “[b]ecause the corporation is a particular type of firm formed by individuals acting voluntarily and for their mutual benefit, it can far more reasonably be viewed as the product of private contract than as a creature of the state” (Fischel 1982, pp. 1273–74). Corporate law should thus provide “off-the-rack” rules that were primarily enabling rather than prescriptive and that could be easily contracted around (Easterbrook and Fischel 1991). Law should not dictate the details of the obligations among the parties because each party is assumed to know her own interests and to protect them best through bargaining and exchange.

Moreover, terms of the corporate “contract,” whether in charters or in state incorporation statutes, were assumed to be correctly priced through an efficient capital market. Because such terms were assumed to have a price associated with them, the complete contract was best seen as consensual in that any shareholder who bought the security could be deemed to have agreed completely to the contract. “All the terms in corporate governance are contractual in the sense that they are fully priced in transactions among the interested parties” (Easterbrook and Fischel 1991, p. 17). Because shareholders could learn about companies from information freely available and could sell stock in a fluid securities market, it was assumed that those who held a company’s stock

voluntarily accepted the risk/return ratio of that security. If the stock tanked, then the shareholder could not blame anyone but herself.

Because the corporation was, in Easterbrook and Fischel's terms, "a voluntary adventure" (1991, p. 12), it became an anachronism to speak of an enforceable fiduciary duty in corporate law. The values of and norms of this version of contractarianism (about which there was disagreement as a matter of contract doctrine) differed "significantly from the values that the fiduciary notion embodie[d] historically and functionally" (Brudney 1997, p. 597). The contractarian move was to characterize fiduciary duties and "its traditional strictures" as a form of a contract, and then "invoke the form as a fulcrum on which to ratchet down the substantive restrictions" (Brudney 1997, p. 597).

The concept of contract and voluntariness was sufficiently dominant that it extended beyond shareholders to all stakeholders of the corporation. Everyone involved in a company—from creditors to employees, from customers to suppliers—was also assumed to be consenting to their involvement. If the parties disliked the terms of the "contract" between themselves and the company, they could leave. Not only could shareholders sell their shares, but employees could quit, managers could find a different company to manage, suppliers could sell their goods elsewhere, and creditors could sell their bonds (Macey 1991).

Even the relationship between firms and the state was seen as voluntary, since companies could easily move their capital and, in the United States, their place of incorporation. If a government jurisdiction sought to impose obligations on firms that went beyond the norm created by competing jurisdictions, companies could leave, creating—depending on who you believed—a regulatory "race to the bottom" (Cary 1974) or an efficiency-creating "race to the top" (Easterbrook and Fischel 1991; Romano 1993). But regardless of the normative characterization, the success of the contractarian mode was clear. Firms did not exist because of a concession from the state; rather firms existed outside the state, and their relationships with government entities (whether nations, states, or communities) were something they opted into and out of.

This left the corporation completely able to disregard any putative obligation that was not explicitly set out in contract, regulation, or common law. According to Fischel (1982), corporations should not be held accountable to stakeholders in ways other than those explicitly set out in the corporate "contract" because doing so would "disrupt the voluntary arrangements that private parties have entered into in forming corporations" (p. 1271). Corporations should not be asked to look after the needs of their employees either, since workers "must look to their contractual rights rather than invoke fiduciary claims" (Easterbrook and Fischel 1991, p. 91). Further, if employees bargained for a certain contract with only limited job security protection (or, more precisely, failed to bargain for a contract that had greater protection), "they ought not grumble if they are held to their bargains when business goes bad. Each investor must live with the structure of risks built into the firm. . . it is all a matter of enforcing the contracts. And for any employee. . . that means the explicit negotiated contract" (Easterbrook and Fischel 1991, p. 37).

Moreover, the absence of protections within the labor contract did not mean, according to contractarians, that workers are unable to bargain for those protections. “Rather, the absence of contractual protections. . . may simply reflect the fact that [nonshareholder] constituencies are unwilling to pay for such protection in the form of lower wages” (Easterbrook and Fischel 1991, p. 36). In other words, if workers wanted to be the beneficiaries of management’s fiduciary duties, they would have simply bargained for such a contractual benefit. Because they had not chosen the protections, they must have chosen the lack of protection. The law should respect their choice.

This notion—whether it went by the name of “choice,” “voluntariness,” or “contract”—was based on a set of beliefs about individual autonomy, rationality, and behavior. The assumption was that people actually chose—which was taken to be a mental decision of some kind usually followed by some kind of physical manifestation of that decision. In the words of Richard Posner, “man is a rational maximizer of his ends in life” (Posner 2003, p. 3). Or in the summary of Anita Bernstein, “This individual knows what he wants and chooses means to reach his goals” (Bernstein 2005, p. 308). But as soon as research accumulated that showed, after genuine inquiries into the substance of choice, that human behavior was much messier and complex than the theory assumed, then it became increasingly clear that the beliefs forming the basis of contractarianism were in fact beliefs. And faith is a poor basis for law.

### 3 FROM LIBERTARIANISM TO AGNOSTICISM IN CORPORATE GOVERNANCE

---

The principal implication of a dedication to rational-choice assumptions in law is that law faces a presumption against its use. A legal regime based on economic rationality will tend to disfavor regulatory efforts, except for those thought necessary to protect the act of efficient choice, such as antifraud initiatives (Friedman 1962). People protect themselves through choice; bad choices are not something with which law should concern itself. Such assumptions “led to conclusions that, more often than not, criticized existing legal constraints as unnecessary or inefficient, and offered a deregulatory normative agenda in place of the legal status quo” (Langevoort 2012, p. 443). With regard to corporate governance, this antiregulatory urge meant that courts looked skeptically on fiduciary duty suits generally, and when reviewing the merit of various suits narrowed the purview and robustness of the traditional notions of the duties of care and loyalty.

As behavioralism has gained traction in the legal academy, the principal implication of its rise is the erosion of this antiregulatory, libertarian presumption (Korobkin 2011). This is not to say that behavioralism necessarily leads to greater governmental intervention, whether by way of regulation or judicial oversight. Rather, in the words of Russell Korobkin, behavioralism replaces the “antiregulation bias” of conventional law and economics with an “agnosticism toward regulation” that allows space for a

“comparative institutional analysis” of whether regulation of various kinds outperforms unregulated markets in different contexts (Korobkin 2011, p. 1659).

Because behavioralism is as messy and indeterminate as the decision-making of the human animal, its prescriptive power will necessarily be less than that of the simple-but-flawed theory of economic rationality. We can be sure that an antiregulatory, libertarian presumption is undeserved, but we are much less sure of the contours of the analytical space now open to us, and what to do within it. As Donald Langevoort has said, “deciding what that intervention should be. . . can itself be vexing” (Langevoort 2011, p. 67).

These problems may be particularly acute in the area of corporate governance for two reasons. First, there is much secrecy and privacy inherent in business decision-making, and it is particularly difficult to observe decision-makers in action. Field studies of decisions in economic settings are relatively few (DellaVigna 2009) and economic studies of boards have often neglected actual board behavior (Ees, Gabriellsson, and Huse 2009). How much economic rationality, or lack thereof, that any one individual or institution will exhibit in any context will depend on a host of factors that will be difficult to know in the abstract. Langevoort again:

The psychologically prudent answer to any question of how someone, for example a CEO in a particular setting or a board of directors involved in group deliberation, will think or act is almost always “it depends,” which does not lend itself to particularly bold or confident legal analysis. (Langevoort 2011, p. 66)

Second, there may be reason to suspect that those who succeed in the business world do so because they are able to be more economically “rational” than the subjects of the myriad behavioral experiments that undermine rationality’s place as a more general predictive device (Tor 2013; Langevoort 2012; Romano 1986). If true, then economic rationality may have more predictive power within corporate analysis than in other areas of the law. And even if this suspicion is only somewhat borne out in experience, the variability of the regulatory outputs of behavioral analysis will be greater in corporate governance than in other areas of law and inquiry.

Langevoort answers this concern by arguing that corporate managers may be subject to greater cognitive errors than the general population with regard to overconfidence, sunk cost bias (an attachment to past decisions, even when wrong), and attribution error (assuming success comes from one’s own doing rather than luck) (Langevoort 2012). For example, because of the nature of competitive tournaments for promotion and financial gain within firms, overconfidence may be “highly adaptive” within the corporate setting (Langevoort 2012, 2007; Tor 2013). Successful risk taking is rewarded, even if *ex ante* the risks are poor ones. And if there are enough people in the tournament, the winner will have been the overconfident one accepting the unreasonable risks. Across all firms the “winners” at the top of the corporate hierarchies are the ones that most exhibit overconfidence and irrational tolerances for risk.

Despite these difficulties, behavioral scholars are making genuine progress in analyzing the workings of business firms and the securities markets, building an extensive

literature. A number of possible regulatory and judicial responses to predictable behavioral phenomena have been identified and described.

One class of behavioral phenomena that has numerous implications for both corporate governance and securities law is the problem of irrational overconfidence (or the less elegantly described overoptimism), which occurs when an individual or a group has a falsely inflated expectation of the chances of success arising from any particular decision. The inflation of expectations can arise from a number of different phenomena. These include the tendency of humans to underestimate the likelihood of small risks, to overestimate the likelihood of salient risks (either because of recency or emotional impact), and to misjudge risks because of confirmation bias (ignoring facts that do not fit with one's previous opinion or worldview, and prioritizing facts that do) (Thaler and Sunstein 2008; Griffin and Brenner 2004; Kahneman, Slovic, and Tversky 1982; Kahneman and Tversky 1973).

An example of these behavioral phenomena that relates to securities regulation pertains to the predominance and persistence of favorable recommendations by sell-side securities analysts (Langevoort 2012; Fisch and Sale 2003). Under traditional, nonbehavioral analysis, this problem is seen as a product of conflict of interest, arising from either collusion between the analysts and the companies whose stock is being pushed, or a desire on the part of analysts to endear themselves with issuers in order to gain their business. The behavioral insight is that instead of (or in addition to) being a product of conflict of interest, the optimism with regard to securities may be the product of natural selection within the industry. The company issuing the security may reward an analyst showing enthusiasm about a particular security with greater access to information, giving the analyst an edge in the marketplace. And if the risks pay off, the analyst's enthusiasm will result in a greater halo effect, which increases the probability that the analyst's favorable recommendations will be followed, which will in turn benefit the company.

If the traditional conflict-of-interest explanation for the predominance of favorable recommendations is more correct, then the law should focus on regulatory responses that depend on disclosure of conflicts and possibly the certification of honesty and good faith on the part of the analyst. A measure of the importance of behavioral research is how poorly these regulatory responses would respond to the problem if based not on conflict but on overoptimism. An overly optimistic analyst may not be motivated by conflicts and is not being dishonest, yet might be wrongly accused of either in a legal regime that presumes such motivations from persistently favorable recommendations (Langevoort 2012). A proper regulatory response would fixate less on honesty and good faith and more on education of all parties of the dangers from, and ubiquity of, overly favorable recommendations.

Overconfidence also skews investment decisions, both by individuals and by firms (Ferris, Jayaraman and Sabherwal 2013; Gervais, Heaton, and Odean 2011; Chuang and Lee 2006). For example, banks with overconfident CEOs take greater risks than their peers, and "top-performing mutual fund managers tend to trade more following their success—to a degree not explained by other factors—and exhibit worse



performance when they do so” (Tor 2013, p. 60). The overconfidence of executives affects both their investment and financing decisions on behalf of their companies (Tor 2013; Malmendier, Tate, and Yan 2007; Paredes 2005). It also “helps explain the volume, type, and financing of mergers and acquisitions activity” and is “linked to aggressive accounting and an increased likelihood of financial misreporting” (Tor 2013, p. 60; see also Roll 1986). Overconfident managers also underestimate the risk of bankruptcy and are overconfident about their ability to avoid liquidation (Dickerson 2003).

Overconfidence, when linked with confirmation bias, can be particularly destructive in the corporate setting (Hamann 2013). Confirmation bias will lead to mistaken views of the firm’s strategic position, and overconfidence will lead managers to be “heavily invested in those beliefs, and hence disinclined to seek out information that would suggest that they might be wrong” (Langevoort 2001, p. 803). Scholars have pointed to a number of famously poor decisions brought about by these skewed perceptions on the part of managers: the failure of the U.S. automobile industry in the 1980s to understand the potential strength of their Japanese competitors (Levinson 1994); the early financial difficulties of Disneyland Paris in the 1990s caused by an overestimation of potential attendance (Hall 2007); fraud allegations at Apple, TimeWarner, and Polaroid (Langevoort 1997); and the general “underperformance of companies undertaking mergers” (DellaVigna 2009, p. 342) caused by the hubris of their executives.

One possible legal implication of these phenomena is that courts should be less eager to depend on the business judgment rule—a presumption of correctness applied to managerial decisions that are informed and not self-interested—in adjudicating claims arising from alleged firm mismanagement. Judicial focus on whether the managerial decision-makers are adequately informed and whether they are burdened by conflicts of interest will not capture defects in decision-making arising from overconfidence (or other biases), even when such defects have potentially disastrous effects on the firm. Behavioral research thus suggests that a more searching inquiry by courts into the substance of business decisions may be appropriate, at least in those instances in which effective (and cognitively unbiased) court scrutiny can counteract defects in decision-making *ex post* or deter them *ex ante*.

Behavioral research has also been very powerful in undermining the notion that securities priced in a fluid market reflect the true, fundamental value of the underlying companies (Stout 2003). This belief in the accuracy of capital market prices underlies many different aspects of corporate and securities law, from the rules of corporate consolidations to the regulatory salience of disclosure. For example, with regard to hostile merger situations, the less confidence one has in market prices the less one can trust individual shareholders to make rational decisions about whether to tender into a hostile bid, and the more likely management on both sides will make mistakes in making or responding to bids (Langevoort 2011). With regard to securities litigation, courts determine liability for alleged frauds in part by deciding whether the untruth was “material” to the typical investor, who is often assumed to be economically rational. The dissonance between how investors actually behave and rationality assumptions



imposed by courts in determining materiality means that courts often fail to protect shareholders from fraud (Hoffman 2006; Huang 2005).

More generally, behavioralists are building a strong case that disclosure is an imperfect and incomplete regulatory prophylactic. Disclosure does not do all the work that rational-choice theory would assume, for a number of reasons. Humans are easily overwhelmed by information, so the benefits of disclosure can be erased with information overload (Paredes 2003; Korobkin and Ulen 2000; chapter 28 by Sunstein in this volume). The amount of data released in routine financial disclosures, for example, can easily swamp most investors' ability to parse the details. Other behavioral phenomena also undermine disclosure as a cure-all. Investors routinely engage in irrational herding behavior in markets, using the decisions of other investors as indicia of value rather than making their own independent judgments based on the disclosed information. (Bainbridge 2000). Irrational risk tolerance is another example of a human bias that, when present, would undermine the power of disclosure as a prophylactic (Choi and Pritchard 2003).

Also, there are data that suggest that disclosure is not only ineffectual but affirmatively harmful in some contexts. Persons who disclose conflicts may feel the disclosure, by warning the counterparties of the risk, has given them moral license to act selfishly (Langevoort 2012; Cain, Loewenstein, and Moore 2005). Indeed, this phenomena is indicative of a larger point about the limits of voluntariness as a legal talisman. If choices are sacrosanct even when constrained or perverse, the powerful party in any exchange is incentivized to manipulate and take advantage of the weaker party (Greenfield 2011a).

Not only is the dependence on disclosure as a legal tool undermined by its limits—for behavioral reasons—in protecting investors against their own mistakes *ex ante* but also by its tendency to be misused—for behavioral reasons—by courts *ex post*. David Hoffman suggests that a legal regime that imposes liability on managers who fail to disclose a material risk is subject to hindsight bias in adjudicating whether the undisclosed matter was material (Hoffman 2006). Hindsight bias is simply the notion that because something has occurred, it was destined to occur. For example, when a judge or jury is faced with the question of whether a fiduciary is liable for failing to disclose the risk of a particular event, such questions will be adjudicated only in those situations in which the event in fact occurred. Because of hindsight bias, the event's occurrence will irrationally influence the adjudicator's estimation of whether the risk should have been disclosed, even though materiality should be evaluated as of the moment the disclosure decision was made. So hindsight bias will tend to lead judges and juries to find materiality—and thus liability for failure to disclose—in situations in which the risk was in fact immaterial. (The same effect has been identified in other *post hoc* evaluations of corporate managers' decisions [Teichman in this volume].)

These examples indicate the profound implications of behavioralism for corporate governance and finance. At the very least, they suggest that the earlier fixation on contractarianism and regulatory libertarianism created perverse effects. They also suggest the benefits of a possible move toward a more robust regime of genuine, enforceable

duties on the part of fiduciaries toward the firm and its investors. At the same time, as exemplified by Hoffman's work, behavioralism can offer insights into the limits of judicial oversight. Because judges are subject to behavioral tendencies and deficiencies as well, there may be instances in which court oversight will not repair or solve the behavioral and cognitive mistakes of corporate managers. This is an area that merits significant additional work: can the oversight of judges, given their own behavioral and cognitive limitations, cause managers to do a better job given theirs? This question is yet to be answered persuasively, which makes Korobkin's assertion of agnosticism quite persuasive, at least at present.

## 4 BEHAVIORALISM AND BOARD MAKEUP

---

In my view, the most potentially profound implications of behavioral research for corporate governance relates to the question of whether the structure and makeup of corporate boards (for example, its group decision-making structure, or the homogeneity of the decision-makers) suppress or inflate irrationalities.

The structure and makeup of boards are quite important for two reasons. First, the fact that the corporation has a sophisticated group decision-maker at the top of the structural hierarchy is seen as a distinctive element of the success of the corporation as a business form (Chandler 1977; Bainbridge 2002; Greenfield 2008b). The board functions as a group decision-maker for the most important questions that the company faces, and in the words of Stephen Bainbridge, "it seems useful to think of the board as a production team" (Bainbridge 2002, p. 8) with decisions being the product produced. When working properly, boards offer material benefits as compared to solitary, individual decision-makers typical in sole proprietorship or to small groups of decision-makers typical in partnerships or similar enterprises. The benefits of group decision-making can be significant and in many cases so outpace individual decision-making that the success of groups is higher not only than the average individual in the group but even higher than the best individual in the group (Bainbridge 2002, p. 8). The relative success of group decision-makers holds both with regard to tasks that have unique correct outcomes and with tasks that require critical evaluative judgment, learning and concept attainment, creativity, and abstract problem solving (Bainbridge 2002, p. 8; Blinder and Morgan 2005; Surowiecki 2004). In complex organizations such as corporations, the benefits of group decision-making may be even more profound, since "the effective oversight of an organization exceeds the capabilities of any individual" and "collective knowledge and deliberation are better suited to this task" (Forbes and Milliken 1999, p. 490).

Second, the form and process of the board is a central issue in corporate law because it is the primary focus of judicial oversight. Courts are typically reluctant to second-guess the substantive business judgment of management, so courts instead evaluate the propriety of challenged managerial judgments on the basis of the process

followed by the board, the information available to it, and its disinterestedness. “[S]o long as a sufficiently independent decision-maker within the approved corporate hierarchy has been given all the material facts and approved some action after due deliberation, the action is beyond serious judicial review” (Langevoort 2012, p. 446).

Traditionally, the disinterestedness obligation of board members was deemed to be violated only when decision-makers had some kind of pecuniary interest at stake, or had a close family member so interested. The worry, of course, was that corporate leaders were acting on the basis of self-interest rather than for the benefit of the corporation. Courts and commentators, therefore, often concentrated on the “independence” of members of the board, as defined by the lack of employment or financial relationship with the company. Decisions made by a board dominated by nonindependent members were seen as tainted; decisions validated by independent members were deemed trustworthy (*Aronson v. Lewis*, 473 A.2d 805 (Del. 1984); *In re Walt Disney Company Derivative Litigation*, 906 A.2d 27 (2006)). The academic literature generally looked at function more than form, but even there the key question was whether board members were sufficiently independent in perspective and attitude to satisfy the board’s monitoring obligation (Brudney 1982; Langevoort 2001).

Especially in the light of the judicial focus on structure and process and on economic ties as sources of disloyalty, behavioral research revealing predictable flaws in group decision-making is likely to pose difficulties for traditional corporate law doctrine. For example, behavioral studies have shown significant skewing effects on decision-making springing from interests other than the mere economic. That is, research is making increasingly clear that the set of motivations that bring into question the independence of fiduciaries extends beyond the pecuniary. One such effect is that of in-group/out-group identification. When those inside a group—board members, for example—are challenged by someone outside the group—a shareholder plaintiff alleging a breach of fiduciary duty, for example—they tend to defend and defer to their fellow group members (Lane et al. 2007; Blair and Stout 2001; Robinson 1996). So even if a group member is independent in that she has no financial interest in the conflict, the psychological tendency to align against the challenger will influence her review of the merits of the challenge. Another relevant effect is the tendency to defer to authority figures (Milgram 1974; Greenfield 2011a). Some have argued that such deference to authority may cause directors to fail effectively to monitor or challenge powerful CEOs (Morck 2008). More generally, any psychological effect that makes it difficult for a director truly to provide an independent check on management is largely ignored by corporate doctrine, even when decision-making is biased significantly (Hill and McDonnell 2007).

In addition, behavioral research has shown predictable defects in decision-making that have less to do with self-interest or disloyalty than with other kinds of bias and mistake. As mentioned above, group decision-making is thought to be a significant reason for the success of corporations as a business form, in part because of a group’s ability to improve on the decision-making of individuals by exposing and mitigating bias and mistake. But these benefits can vanish, and indeed transform into costs, if the group

reinforces bias and submerges mistakes, worsening irrationalities. As Avishalom Tor has recently written, “the evidence shows small-groups outperform individual rationality in some cases but at other times exhibit similar or even more extreme judgmental biases and decision errors” (Tor 2013, p. 63).

“Groupthink” is the most common example of this phenomenon, and is particularly relevant for corporate governance. “Groupthink” is the label attached to mistakes made by institutional decision-makers when the presence of similarly thinking participants in a group results in biases being reinforced rather than challenged, and mistakes validated rather than exposed (Janis 1983; O’Connor 2003; Hall 2007). Another example of group tendencies that worsen decision-making is the inclination for discussion within groups comprised of individuals with similar worldviews and perspectives to harden those perspectives and views. In discussions about political issues, for example, groups on the extremes of political discourse become more extreme after discussion within the group (Isenberg 1986; Sunstein 2005; Tor 2013). These implications are greater within groups that are homogeneous in perspective and in racial, gender, and class composition, since “defective decisionmaking” is “strongly correlated” with structural flaws such as “insulation and homogeneity” (Sunstein 2003). As Jolls and Sunstein have articulated, “erroneous judgments often result when deliberations are undertaken by like-minded people” (Jolls and Sunstein 2006, p. 218).

The worry from a corporate governance perspective is that the quality of the decision-making of the board is eroded when its homogeneity and insularity make it less likely that ideas will be properly vetted or assumptions appropriately challenged (Greenfield 2008b). If a key element of the success of the corporation as a business form is the presence of a sophisticated group decision-maker at the top of the business structure, this success is put at risk when the board suffers from structural or formational defects that weaken the decisional process or skews the results.

Like the worry about bias and disloyalty, this concern about the insularity and homogeneity of the board is not mitigated by the traditional corporate law insistence on board member independence. The mere fact that a board member is not employed by the company does not correlate well with the needed diversity of perspective. Instead, attention needs to turn to the actual makeup of the board, with the goal of using greater board pluralism as a tool to create a board culture that encourages dissent and challenge.

If the homogeneity of groups is a reason to worry about the quality of its decisions, then the current makeup of most boards is quite flawed. In fact, corporate boards may be the least diverse powerful institutions in the United States. Scholars increasingly point out the gender and racial homogeneity of boards and executive suites, and the dangers to decision-making posed by such narrowness (Fairfax 2011; Branson 2010; Broome and Krawiec 2008; Branson 2007; Carbado and Gulati 2004). As recently as 1996, no woman had ever been the CEO of a Fortune 500 company in the United States. Even more recently, women held less than 3% of the CEO positions in the Fortune 1000 in 2010 (Branson 2010). As of 2012, only six Fortune 500 chief executives are African

American, seven are Asian, and six are Latino (Black Entrepreneur Profiles 2012; DiversityInc 2011).

The numbers with regard to board membership are no better. As of 2006, 51.2% of US corporations had no or only one female director (Branson 2007). Recent data show that less than 17% of Fortune 500 board seats are held by women (Catalyst.org 2011), and this number inflates the level of true diversity of perspective by double counting women who sit on several boards (Branson 2010). In the Fortune 100, men held 82% of board seats in 2010; minority men and women together only held approximately 15% of seats (Alliance for Board Diversity 2011).

These numbers look particularly poor when considered alongside the research showing the necessity for a “critical mass” of minority group members to be present before a group’s dynamics change (Konrad, Kramer, and Erkut 2008; Kramer, Konrad, and Erkut 2006). Indeed, some research indicates that the presence of a minority whose numbers fall below the critical mass are perceived as “tokens,” making the dominant group cohere more, isolating the minority from the rest of the group (Elstad and Ladegard 2010; Gustafson 2008; Robertson and Park 2007; Kanter 1977). These studies raise the possibility that in the corporate context, boards with some level of diversity but that nevertheless fail to reach a “critical mass” will not enjoy the potential benefits of diversity and may still fall victim to the decision-making flaws that spring from homogeneity.

In the United States, data about the lack of board diversity are used primarily as part of fairly gentle efforts to encourage companies to diversify their leadership. In Europe, on the other hand, governments have begun to pressure companies in various ways to improve board diversity. France, Spain, and Norway, for example, now require companies to reserve 40% of board seats for women. The UK government recommends (but does not require) that women make up at least a quarter of large company boards (Werdigier 2011; Elstad and Ladegard 2010).

With these data in mind, the value of more diverse boards is an area where behavioral research is likely to be quite helpful going forward. If diversity does indeed assist in debiasing boards, then more diverse boards should experience improvements in their decision-making. While the question of how to measure such improvements is tricky indeed (especially given the need to distinguish between the direct effects associated with adding new groups such as women to boards, and the secondary effects associated with diversifying the composition of boards), behavioral research is beginning to amass a fairly convincing set of studies that bolster the hypothesis that diversity improves board (as opposed to company) performance. For example, boards with more women have been shown to be more active and independent in monitoring management, to be more likely to engage with the company’s stakeholders, to show more attention to risk oversight and control, and to be more likely to be concerned about social responsibility (Brown, Brown, and Anastopoulos 2002). There is also evidence that the presence of women on a board improves the quality of board deliberations (Burke and Vinnicombe 2008; Robertson and Park 2007; Huse and Solberg 2006; Van der Walt and Ingleby 2003), in part by empowering “constructive dissent” that can lead

to “board unity,” which is “essential to setting a clear strategic direction and to overseeing risk and resources” (Brown, Brown, and Anastopoulos 2002, p. 5). One notable finding is that male directors attend more board meetings when the board is more gender diverse (Adams and Ferreira 2009). Other recent findings indicate that more diverse business teams are able to take advantage of a wider pool of relevant knowledge (Hoogendoorn and van Praag 2012). In the words of Aaron Dhir, “establishing a level of ‘cognitive diversity’ in the boardroom is... a key strategic asset which serves to assist the firm in averting the perils and docile conduct associated with groupthink” (Dhir 2010, p. 595).

Two caveats to this body of research need articulation. The first is that any benefits to diversity will have to be measured against the possible costs of diversity, including a less streamlined decision-making process and less intragroup trust, at least initially (Langevoort 2001). From the standpoint of institutional design (as opposed to justice, for example) the goal is to have boards with “enough diversity to encourage the sharing of information and active consideration of alternatives, but enough collegiality to sustain mutual commitment and make consensus-reaching practicable” (Langevoort 2001, pp. 810–11).

The other caveat has to do with the link between board performance and company financial performance. Assuming board management matters, we should expect to see improvements in board decision-making manifested in improvement in company financial performance. To date, however, the evidence as to whether diversity of boards leads to measurable improvements in company financial performance is largely equivocal (Carter et al. 2010; Dhir 2010; Branson 2007; Robertson and Park 2007; Farrell and Hersch 2005). Dhir, for example, reports that while some studies find a correlation between board diversity and profitability, the causal link is unclear and its direction uncertain (Dhir 2010). And even when a causal link is indicated, the data suggest that some board functions benefit from diversity while others do not (Dhir 2010).

Of course these equivocal data do not disprove the arguments in favor of board diversity. The data might be unclear because the correct “mix” of diversity has yet to be calibrated, or it could suggest that the benefits of diversity to board performance accrue in ways that do not appear in the financial data in the short term. It is also important to note that evaluating the benefits of diversity only in terms of financial data ignores diversity’s nonfinancial benefits and undervalues its other rationales (Fairfax 2011; Joo 2004).

All in all, the research remains undeveloped enough that it is yet unknown how best to create the conditions whereby the benefits of diversity are maximized and the costs minimized. It seems clear enough that sufficient numbers are necessary to overcome the problems of tokenism and self-censorship. Efforts to clarify how best to cure the problems of homogeneity are almost certainly to be worth the scholarly and regulatory attention.

Finally, it is also worth considering that gender and racial diversities are hardly the only kinds of pluralism likely to be effectual in debiasing board decisions and protecting against groupthink. In fact, class differences may trump racial and gender



differences as proxies for distinctiveness of perspective (Fairfax 2005). I have argued elsewhere that broadening the board (even by way of mandatory rule) to include directors elected by employees and other stakeholders would improve board deliberation (Greenfield 2006, 2008a, 2008b, 2012), and others have made analogous arguments for other stakeholders (Yosifon 2009). There is firm-level and country-level data that can be cited in support of “co-determined” boards, if not for stakeholder boards (Cochon 2011). Some business leaders, for example, have suggested that the resilience of Germany in the face of the global financial crisis is owing, at least in part, to German companies’ inclusion of employee representatives on supervisory boards (Cochon 2011). But behavioral research is quite thin in this area, with insights coming from extrapolating from behavioral research regarding political differences among members of groups, the creation of in-group identity, and the salience of the tendency for reciprocity (Greenfield 2006).

As behavioral scholarship on diversity grows in influence and scope, it is worth emphasizing that traditional corporate law jurisprudence is unlikely to be sufficiently nimble or accommodating to do much about its insights. The board’s own makeup falls squarely within the sphere of business judgment toward which US courts generally genuflect. As of 2010, the Securities and Exchange Commission does require US companies to disclose “whether, and if so how, the nominating committee. . . considers diversity in identifying nominees for director” (Fairfax 2011), but there is no teeth in the provision and the kinds of diversity at issue are few. The Sarbanes-Oxley Act of 2002 incentivized greater board member independence, especially on auditing committees, which some have argued will help with the debiasing project (Jolls and Sunstein 2006). But as mentioned above, behavioral research suggests that the independence called for is not the kind that will lead to real diversity of perspectives (Marnet 2008). All in all, going forward, there is certainly little jurisprudential traction for broadened fiduciary duties that would include a nod toward an obligation of greater diversity of various kinds. Attention to these issues will depend instead on shareholder activists, enlightened management teams, and further activity by legislators and regulators.

## 5 CONCLUSION

---

In crafting public policy, accurate predictions of human behavior are essential in anticipating the effects of various regulatory options. In adjudicating disputes, accurate descriptions of motivations and understandings on which past behavior was based are similarly essential in allocating liability and deciding remedy. So whether predicting future behavior or judging past behavior, an understanding of human decision-making is crucial. The implications of behavioral research are thus significant for many areas of public policy and law, since it challenges and upsets the traditional notions of economic rationality that have long served as touchstone assumptions on which these predictions and descriptions were based.



In corporate and securities law, these implications are likely to be even more profound. Beliefs about the unbounded economic rationality of participants in the “nexus of contracts” that was assumed to embody the modern corporation animated views about the proper role of courts, the strength of fiduciary duty, the efficiency of securities markets, the power of disclosure, the obligations toward stakeholders, and even whether law itself could be contracted around. So as behavioral research increasingly weakens these assumptions, many areas of corporate policy and doctrine will require rethinking. Researchers and other academics are further along in this project than are courts, but much work remains on all fronts.

## ACKNOWLEDGMENTS

---

I thank Eyal Zamir, Doron Teichman, and participants of the conference behavioral law and economics at Notre Dame Law School for their helpful comments. This paper was supported by the Dean’s Research Fund at Boston College Law School.

## REFERENCES

---

- Adams, R. B., and D. Ferreira. 2009. Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics*, 94: 291–309.
- Andrews, E. L. 2008. Greenspan concedes error on regulation. *New York Times*, October 23.
- Alliance for Board Diversity. 2011. *Missing Pieces: Women and Minorities on Fortune 500 boards, 2010 Alliance for Board Diversity Census*. Available at [http://theabd.org/ABD\\_report.pdf](http://theabd.org/ABD_report.pdf).
- Arlen, J., M. Spitzer, and E. Talley. 2002. Endowment effects within corporate agency relationships. *Journal of Legal Studies*, 31: 1–37.
- Bainbridge, S. M. 2000. Mandatory disclosure: A behavioral analysis. *University of Cincinnati Law Review*, 68: 1023–60.
- Bainbridge, S. M. 2002. Why a board? Group decisionmaking in corporate governance. *Vanderbilt Law Review*, 55: 1–55.
- Bebchuk, L., A. Cohen, and A. Ferrell. 2002. Does the evidence favor state competition in corporate law? *California Law Review*, 90: 1775–821.
- Bernstein, A. 2005. Whatever happened to law and economics? *Maryland Law Review*, 64: 101–303.
- Black Entrepreneur Profiles. 2012. African American CEO’s of Fortune 500 companies. Retrieved from <http://www.blackentrepreneurprofile.com/fortune-500-ceos/>.
- Blinder, A. S., and J. Morgan. 2005. Are two heads better than one? Monetary policy by committee. *Journal of Money, Credit and Banking*, 37: 789–811.
- Blair, M., and L. Stout. 2001. Trust, trustworthiness, and the behavioral foundations of corporate law. *University of Pennsylvania Law Review*, 149: 1753–810.
- Branson, D. M. 2007. *No Seat at the Table: How Corporate Governance and Law Keep Women Out of the Boardroom*. New York: NYU Press.

- Branson, D. M. 2010. *The Last Male Bastion: Gender and the CEO Suite at Public Companies*, D. Branson, ed. New York: Routledge.
- Broome, L. L., and K. D. Krawiec. 2008. Signaling through board diversity: Is anyone listening? *University of Cincinnati Law Review*, 77: 431–72.
- Brown, D. A. H., D. Brown, and V. Anastosopoulos. 2002. *Women on Boards: Not Just the Right Thing... But the "Bright" Thing*. Ottawa: Conference Board of Canada.
- Brudney, V. 1982. The independent director: Heavenly city or Potemkin village? *Harvard Law Review*, 95: 597–659.
- Brudney, V. 1997. Contract and fiduciary duty in corporate law. *Boston College Law Review*, 38: 595–665.
- Burke, R. J., and S. Vinnicombe. 2008. Women on corporate boards of directors: International issues and opportunities. In *Women on Corporate Boards of Directors: International Research and Practice*, S. Vinnicombe, V. Singh, R. J. Burke, D. Bilimoria, and M. Huse, eds. Northampton: Edward Elgar, 1–11.
- Cain, D., G. Loewenstein, and D. Moore. 2005. The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *Journal of Legal Studies*, 34: 1–25.
- Carbado, D. W., and G. M. Gulati. 2004. Race to the top of the corporate ladder: What minorities do when they get there. *Washington and Lee Law Review*, 61: 1643–91.
- Carter, D. A., F. D'Souza, B. J. Simpkins, and W. G. Simpson. 2010. The gender and ethnic diversity of US boards and board committees and firm financial performance. *Corporate Governance: An International Review*, 18(5): 396–414.
- Cary, W. L. 1974. Federalism and corporate law: Reflections upon Delaware. *Yale Law Journal*, 83: 663–705.
- Catalyst.org. 2011. Fortune 500 board seats held by women. Retrieved from <http://www.catalyst.org/knowledge/fortune-500-board-seats-held-women>.
- Chandler, A. 1977. *The Visible Hand: The Managerial Revolution in American Business*. Cambridge, MA: Harvard University Press.
- Choi, S., and A. Pritchard. 2003. Behavioral economics and the SEC. *Stanford Law Review*, 56: 1–73.
- Chuang, W., and B. Lee. 2006. An empirical evaluation of the overconfidence hypothesis. *Journal of Banking and Finance*, 30: 2489–515.
- Cochon, A. 2011. *Board Level Employee Representation Rights in Europe*. Brussels: European Trade Union Institute.
- DellaVigna, S. 2009. Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47: 315–72.
- Dhir, A. A. 2010. Towards a race and gender-conscious conception of the firm: Canadian corporate governance, law and diversity. *Queen's Law Journal* 35: 569–624.
- Dickerson, A. M. 2003. Behavioral approach to analyzing corporate failures. *Wake Forest Law Review*, 38: 1–54.
- DiversityInc. 2011. Where's the diversity in Fortune 500 CEOs? Retrieved from <http://www.diversityinc.com/facts/wheres-the-diversity-in-fortune-500-ceos/>.
- Easterbrook, F. H., and D. R. Fischel. 1991. *The Economic Structure of Corporate Law*. Cambridge, MA: Harvard University Press.
- Ees, H. V., J. Gabriëlsson, and M. Huse. 2009. Toward a behavioural theory of boards and corporate governance. *Corporate Governance: An International Review*, 17: 307–19.
- Elstad, B., and G. Ladegard. 2010. Women on corporate boards: Key influencers or tokens? *Journal of Management and Governance*, 16: 595–615.

- Fairfax, L. M. 2005. The bottom line on board diversity: A cost-benefit analysis of the business rationales for diversity on corporate boards. *Wisconsin Law Review*, 2005: 795–854.
- Fairfax, L. M. 2011. Board diversity revisited: New rationale, same old story? *North Carolina Law Review*, 89: 855–85.
- Farrell, K. A., and P. L. Hersch. 2005. Additions to corporate boards: The effect of gender. *Journal of Corporate Finance*, 11: 85–106.
- Ferris, S. P., N. Jayaraman, and S. Sabherwal. 2013. CEO overconfidence and international merger and acquisition activity. *Journal of Financial and Quantitative Analysis*, 48: 137–64.
- Fisch, J., and H. Sale. 2003. The securities analyst as agent: Rethinking the regulation of analysts. *Iowa Law Review*, 88: 1035–97.
- Fischel, D. R. 1982. The corporate governance movement. *Vanderbilt Law Review*, 35: 1273–74.
- Forbes, D. P., and F. J. Milliken. 1999. Cognition and corporate governance: Understanding boards of directors as strategic decision-making groups. *Academy of Management Review*, 24: 489–505.
- Friedman, M. 1962. *Capitalism and Freedom*. Chicago: University of Chicago Press.
- Gervais, S., J. B. Heaton, and T. Odean. 2011. Overconfidence, compensation contracts, and capital budgeting. *Journal of Finance*, 66: 1735–77.
- Greenfield, K. 2002. Using behavioral economics to show the power and efficiency of corporate law as regulatory tool. *University of California Davis Law Review* 35: 581–644.
- Greenfield, K. 2006. *The Failure of Corporate Law: Fundamental Flaws and Progressive Possibilities*. Chicago: University of Chicago Press.
- Greenfield, K. 2008a. Defending stakeholder governance. *Case Western Reserve Law Review*, 58: 1043–65.
- Greenfield, K. 2008b. Saving the world with corporate law. *Emory Law Journal*, 57: 948–84.
- Greenfield, K. 2011a. *The Myth of Choice: Personal Responsibility in a World of Limits*. New Haven: Yale University Press.
- Greenfield, K. 2011b. Unconscionability and consent in corporate law (a comment on Cunningham). *Iowa Law Review Bulletin*, 96: 92–99.
- Greenfield, K. 2012. The stakeholder strategy. *Democracy: A Journal of Ideas*, 26: 47–59.
- Griffin, D., and L. Brenner. 2004. Perspective on probability judgment calibration. In *Blackwell Handbook of Judgment and Decision Making*, D. J. Koehler and N. Harvey, eds. Malden, MA: Blackwell, 177–99.
- Gustafson, J. L. 2008. Tokenism in policing: An empirical test of Kanter’s hypothesis. *Journal of Criminal Justice*, 36: 1–10.
- Hall, K. 2007. Looking beneath the surface: The impact of psychology on corporate decision making. *Managerial Law*, 49: 93–105.
- Hamann, H. 2013. *Unpacking the Board: A Comparative and Empirical Perspective on Groups in Corporate Decision-Making*. Book manuscript, on file with author.
- Hill, C. A., and B. H. McDonnell. 2007. Disney, good faith, and structural bias. *Journal of Corporation Law*, 32: 833–64.
- Hoffman, D. A. 2006. The “Duty” to Be a Rational Shareholder. *Minnesota Law Review*, 90: 537–611.
- Hoogendoorn, S., and M. van Praag. 2012. Ethnic performance and team performance: A field experiment. Discussion Paper 6731, Institute for the Study of Labor (Germany), July.
- Huang, P. H. 2005. Moody investing and the Supreme Court: Rethinking the materiality of information and the reasonableness of investors. *Supreme Court Economic Review*, 13: 99–111.

- Huse, M., and A. G. Solberg. 2006. Gender-related boardroom dynamics: How Scandinavian women make and can make contributions on corporate boards. *Women in Management Review*, 21(2): 113–30.
- Isenberg, D. J. 1986. Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, 50: 1141–51.
- Janis, I. L. 1983. *Groupthink*. 2nd ed. Boston: Houghton-Mifflin.
- Jensen, M. C., and W. H. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3: 305–60.
- Jolls, C., and C.R. Sunstein. 2006. Debiasing through law. *Journal of Legal Studies*, 35: 199–241.
- Jolls, C., C. R. Sunstein, and R. Thaler. 1998. A behavioral approach to law and economics. *Stanford Law Review*, 50: 1471–550.
- Joo, T. W. 2004. Race, corporate law, and shareholder value. *Journal of Legal Education*, 54: 351–64.
- Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- Kahneman, D. and A. Tversky. 1973. On the psychology of prediction. *Psychological Review*, 80: 237–51.
- Kanter, R. M. 1977. *Men and Women of the Corporation*. New York: Basic Books.
- Konrad, A. M., V. Kramer, and S. Erkut. 2008. The impact of three or more women on corporate boards. *Organizational Dynamics*, 37: 145–164.
- Korobkin, R. B. 2011. What comes after victory for behavioral law and economics? *Illinois Law Review*, 2011: 1653–74.
- Korobkin, R. B., and T. S. Ulen. 2000. Law and behavioral science: Removing the rationality assumption from law and economics. *California Law Review*, 88: 1051–144.
- Kramer, V. W., A. M. Konrad, and S. Erkut. 2006. *Critical Mass on Corporate Boards: Why Three or More Women Enhance Governance*. Wellesley, MA: Wellesley Centers for Women.
- Lane, K. A., et al. 2007. Implicit social cognition and law. *Annual Review of Law and Social Science*, 3: 427–51.
- Langevoort, D. 1997. Organized illusions: A behavioral theory of why corporations mislead stock market investors (and cause other social harms). *University of Pennsylvania Law Review*, 146: 101–72.
- Langevoort, D. 1998. Behavioral theories of judgment and decisionmaking in legal scholarship: A literature review. *Vanderbilt Law Review*, 51: 1499–528.
- Langevoort, D. 2000. Organized illusions: A behavioral theory of why corporations mislead stock market investors (and cause other social harms). In *Behavioral Law and Economics*, C. Sunstein, ed. Cambridge: Cambridge University Press, 144–67.
- Langevoort, D. 2001. The human nature of corporate boards: Law, norms, and the unintended consequences of independence and accountability. *Georgetown Law Journal*, 89: 797–832.
- Langevoort, D. 2007. Diversity and discrimination from a corporate perspective: Grease, grit and the personality types of tournament survivors. In *NYU Selected Essays on Labor and Employment Law: Behavioral Analyses of Workplace Discrimination*, vol. 3, M. Gulati and M. Yelnosky, eds. New York: Aspen, 141–62.
- Langevoort, D. 2011. The behavioral economics of mergers and acquisitions. *Transactions: Tennessee Journal of Business Law*, 12: 65–79.
- Langevoort, D. 2012. Behavioral approaches to corporate law. In *Research Handbook on the Economics of Corporate Law*, C. A. Hil and B. H. McDonnell, eds. Northampton, MA: Edward Elgar, 442–55.

- Levinson, H. 1994. Why the behemoths fell: Psychological roots of corporate failure. *American Psychologist*, 49: 428–36.
- Macey, J. R. 1991. An economic analysis of the various rationales for making shareholders the exclusive beneficiaries of corporate fiduciary duties. *Stetson Law Review* 21: 23–44.
- Malmendier, U., G. Tate, and J. Yan. 2007. Corporate financial policies with overconfident managers. National Bureau of Economic Research Working Paper no. 13570.
- Marnet, Oliver. 2008. Behaviour and rationality in corporate governance. *International Journal of Behavioural Accounting and Finance*, 1: 4–22.
- Milgram, S. 1974. *Obedience to Authority: An Experimental View*. New York: Harper and Row.
- Morck, R. 2008. Behavioral finance in corporate governance: Economics and ethics of the devil's advocate. *Journal of Management and Governance*, 12: 179–200.
- O'Connor, M. 2003. The Enron board: The perils of groupthink. *University of Cincinnati Law Review*, 71: 1233–319.
- Paredes, T. A. 2003. Blinded by the light: Information overload and its consequences for securities regulation. *Washington University Law Quarterly*, 81: 417–86.
- Paredes, T. A. 2005. Too much pay, too much deference: Behavioral corporate finance, CEOs, and corporate governance. American Law and Economics Association Annual Meeting.
- Posner, R. A. 2003. *Economic Analysis of Law*. 6th ed. New York: Aspen.
- Robertson, Q. M., and H. J. Park. 2007. Examining the link between diversity and firm performance: The effects of diversity reputation and leader racial diversity. *Group and Organization Management*, 32: 548–68.
- Robinson, W. P., ed. 1996. *Social Groups and Identities: Developing the Legacy of Henri Tajfel*. Oxford: Butterworth-Heinemann.
- Roll, R. 1986. The hubris hypothesis of corporate takeovers. *Journal of Business*, 59: 197–216.
- Romano, R. 1986. A comment on information overload, cognitive illusions and their implications for public policy. *Southern California Law Review*, 59: 313–28.
- Romano, R. 1993. *The Genius of American Corporate Law*. Washington, DC: American Enterprise Institute.
- Stout, L. A. 2003. The mechanisms of market inefficiency: An introduction to the new finance. *Journal of Corporate Law*, 28: 63–69.
- Sunstein, C. R. 2003. *Why Societies Need Dissent*. Cambridge, MA: Harvard University Press.
- Sunstein, C. R. 2005. Group judgments: Deliberation, statistical means, and information markets. *New York University Law Review*, 80: 962–1049.
- Surowiecki, J. 2004. *The Wisdom of Crowds*. New York: Doubleday.
- Thaler, R., and C. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.
- Tor, A. 2014. Understanding Behavioral antitrust. *Texas Law Review*, 92: 573–667.
- Van der Walt, N., and C. Ingley. 2003. Board dynamics and the influence of professional background, gender, and ethnic diversity of directors. *Corporate Governance*. 15: 291–334.
- Werdigier, J. 2011. Britain joins calls to close boardroom gender gap. *International Herald Tribune*, February 25.
- Yosifon, D. 2009. The consumer interest in corporate law. *University of California Davis Law Review*, 43: 253–313.

## CHAPTER 21

---

# THE MARKET, THE FIRM, AND BEHAVIORAL ANTITRUST

---

AVISHALOM TOR

## 1 INTRODUCTION

---

ALTHOUGH one legal scholar considered the implications of a psychological phenomenon for antitrust law already in the 1980s (Gerla 1985), the more systematic development of a behavioral approach in this field dates back only to the turn of the twenty-first century (Aviram and Tor 2004; Tor 2002, 2003, 2004), some years after behavioral law and economics already had made significant inroads in many other legal fields (Jolls, Sunstein, and Thaler 1998; Langevoort 1998). Even then, perhaps due to the dominance of rationality-based law and economics in antitrust, the behavioral approach took significant additional time to garner broad attention in the field (Leslie 2010). Yet behavioral antitrust has become increasingly popular and hotly debated more recently, once commentators began recognizing the inevitable and significant antitrust challenge posed by robust findings of systematic and predictable deviations from strict rationality (Stucke 2010).

Some enthusiastic proponents of behavioral antitrust depict it as a wholesale alternative to traditional antitrust law and economics (e.g., Horton 2011; Stucke 2013), while vocal detractors criticize the approach on numerous grounds (Wright and Stone 2012). A closer examination reveals, however, that both extreme positions in the behavioral antitrust debate are mistaken. Thus a better understanding of antitrust actors' behavior indeed can advance policy and doctrine, but does not offer a complete substitute for the accepted economic analysis of antitrust law. Moreover, beyond the familiar arguments surrounding the application of behavioral evidence to the law more generally (Mitchell 2002; Tor 2008), behavioral antitrust faces an additional set of external validity challenges, shared by only a few other legal fields (Arlen 1998). While antitrust law primarily addresses the behavior of firms in market settings, much of the empirical



evidence that the behavioral approach draws on concerns individual behavior outside the specific institutional environment of the firm, often in nonmarket settings.

The centrality of markets and firms in antitrust means that behaviorally informed analyses of antitrust law must account for the effects of these institutions on the rationality of antitrust actors. Yet in practice scholars frequently either ignore such institutional effects altogether (Huffman 2012) or simply assume, without more, that they guarantee the rationality of all the antitrust-relevant conduct of firms in markets (Werden, Froeb, and Shor 2011, Wright and Stone 2012). In reality, however, markets and firms facilitate rationality in many circumstances but inhibit or fail to promote it on other occasions. Hence both commentators who disregard their powerful effects altogether and those who unquestioningly rely on firm and market institutions always to produce rational behavior inevitably reach some erroneous antitrust conclusions.

This chapter begins by explaining the centrality of hypothetical, strict rationality in antitrust, then examines the complex interplay of markets, firms, and the degree of rationality manifested by real antitrust actors. The implications of this interplay for antitrust doctrine and policy in areas ranging from horizontal and vertical restraints of trade, through monopolization, to merger enforcement practices follows. The chapter concludes by outlining some important open questions and future research directions for antitrust concerning of firms, markets, and rationality.

## 2 ANTITRUST AND RATIONALITY

---

Present day antitrust—perhaps more than any other legal field—is based on the traditional economic assumption that all market participants are rational decision-makers. The firms whose market behavior is the focus of the field are assumed to be perfectly rational competitors that make strictly rational judgments and whose decisions seek always and only to maximize profits (Areeda and Hovenkamp 2006). Moreover, the microeconomic model of competition that the law relies on assumes that consumers are rational actors as well (Werden and Froeb 2008).

The rationality assumption also has concrete legal manifestations throughout antitrust doctrine and enforcement policy. In the United States, for example, the Supreme Court made the legal standard for allegations of illegal monopolization by predatory pricing under Section 2 of the Sherman Act nearly insurmountable by relying on the rationality assumption in *Brooke Group Ltd. v. Brown & Williamson Tobacco Corp.*, 519 U.S. 209 (1993). The Court declared that conduct will not amount to predatory pricing unless the alleged scheme involved pricing below some measure of cost and the predator had a rational prospect of recouping its losses. *Brooke Group* then concluded that predatory pricing schemes only rarely are tried and even more rarely are successful. According to this view, for recoupment to be likely the predator *inter alia* must have a very large market share that is protected by significant entry barriers. However,



because few alleged predators meet the former condition and few markets the latter, *Brooke Group* concluded that price predation rarely occurs. Hence the Court declared that predatory pricing allegations can be rejected summarily in the common case of unlikely recoupment. The same rationale was applied by the Court more recently in *Weyerhaeuser Co. v. Ross-Simmons Hardwood Lumber Co.*, 549 U.S. 312 (2007), to reject allegations of predatory bidding, because “a rational firm would not willingly suffer definite, short-run losses” without “a reasonable expectation” of recoupment. More generally, the *Weyerhaeuser* Court noted that a “rational business will rarely make th[e] sacrifice” involved in such predation (*id.*, p. 323).

Importantly, the Court’s reliance on the rationality assumption to shape antitrust doctrine is not limited to Section 2 predation. A few years prior to *Brooke Group*, for instance, the Court in *Matsushita Electric Industrial Co. v. Zenith Radio Corp.*, 475 U.S. 574 (1986) summarily rejected allegations of a predatory horizontal conspiracy once it determined that maintenance of the conspiracy would have required the alleged conspirators to behave irrationally according to traditional economic models. Assumptions of rationality on the part of firms and consumers alike also have played a role in the Court’s Section 1 jurisprudence with respect to vertical restraints between manufacturers and their distributors—such as tying arrangements and resale price maintenance (Tor and Rinner 2011)—and impact antitrust enforcement when the agencies evaluate whether proposed mergers are likely substantially to lessen competition under Section 7 of the Clayton Act (Werden and Froeb 2008).

Models based on assumptions of strict rationality clearly are pervasive in antitrust law, shaping a variety of doctrines across the field and playing a role in merger policy. The extensive behavioral evidence of bounded rationality therefore poses an obvious challenge to present-day antitrust, suggesting that some doctrines and enforcement practices may require reevaluation, possibly revision. Yet before calling upon antitrust to account for bounded rationality, the behavioral approach must address a basic question of external validity (Tor 2008). After all, the bulk of the behavioral evidence concerns individuals and often involves nonmarket behavior. Potentially, therefore, the otherwise extensive evidence of systematic and predictable deviations from standard rationality on the part of individuals may not apply to the conduct of firms and their consumers in market settings.

The response of behavioral antitrust to these external validity concerns is twofold. Analysts should draw as much as possible on both experimental and observational studies of firm and market behavior, thereby reducing the need to extrapolate from more general behavioral findings. More direct, quantitative evidence of antitrust-relevant behavior on the part of firms is extremely limited, however. Antitrust scholars must therefore also judge the external validity of extant behavioral findings by carefully examining the processes through which markets and firms variously promote and inhibit rationality and the implications of these processes for antitrust analysis, a task to which we now turn.

## 3 THE MARKET

---

Markets often promote more rational behavior by consumers and firms, but can also allow and even facilitate some forms of bounded rationality.

### 3.1 Consumers in the Market

When markets offer good information, consumers' judgments and decisions may be more accurate and better aligned with their preferences than in nonmarket settings. The available evidence, however, paints a complex picture. For one, the products and services that consumers must choose among do not always justify a commitment of significant time, cognitive, or financial resources to make optimal judgments and decisions, so consumers rationally ignore some relevant information. Producers who expect to benefit from consumers' educated choices may respond by providing relevant information to consumers via advertising campaigns, marketing, and similar efforts. Such responses not only tap the superior information that producers already possess about their products and services, but also offer significant economies of scale, given the low cost of offering similar information to many consumers (Stigler 1961). Nevertheless, insofar as numerous competing producers offer such information, consumers still must determine which products and services best match their preferences.

Moreover, despite the increasing abundance of information—and occasionally because of it—many consumers still commonly and routinely make product and service choices that are suboptimal for them. Even when competition is present, producers in some markets prefer to offer only partial or opaque information to limit the ability of consumers to evaluate their products. Specifically, producers can benefit by designing products that lead more naive consumers to make inferior, costly decisions—as in the case of some credit card plans—that both increase producers' profits and subsidize the superior products chosen by more sophisticated consumers, helping attract the latter as well (Gabaix and Laibson 2006). In other instances, firms may develop products that are more complex than necessary—such as certain cellular service plans—making it exceedingly difficult to compare their offerings (Ayal 2011; Bar-Gill 2012a).

Markets thus often provide consumers abundant information that can facilitate better judgments and decisions, but consumers still face significant challenges. Insofar as the interests of producers and consumers are not fully aligned, the latter frequently are at a fundamental disadvantage compared to the former—who have the experience, opportunity, and resources needed to exploit at least some consumers.

Besides the role of information, one familiar argument is that markets make deviations from rationality irrelevant because they cancel out in the aggregate so that markets overall perform as if they were comprised of rational participants (Friedman 1953).

This argument, however, fails to account for *systematic* deviations from rationality that bias market behavior in predictable and consistent directions and therefore do not cancel out in the aggregate (Tor 2002).

A different argument asserts that one can derive the main implication of traditional economic models of consumer behavior—namely, the negatively sloping demand curve that associates higher prices with lower demand—without assuming rationality, only that consumers have some resource constraint (Becker 1962). Yet this insight is not particularly helpful for antitrust law, whose doctrine and merger enforcement policies both rely on assumptions of consumer rationality well beyond setting up negatively sloping demand curves (Bennett et al. 2010; Werden, Froeb, and Shor 2011). For example, antitrust analyses that ignore the “sticky” behavior of real-world consumers—who are more reluctant to switch among products than rationality-based models assume—may arrive at excessively broad market definitions that understate the market shares of merging firms.

## 3.2 Producers in the Market

Beyond providing them with incentives and opportunities to react to the bounded rationality of consumers, markets also help align producers’ own behavior with rationality through a number of mechanisms.

### 3.2.1 Aggregation Mechanisms

Economists have long argued that markets overall may comport with the predictions of strictly rational models even while individual firms deviate from it when random errors cancel out in the aggregate. Nonetheless, we noted already that systematic deviations from rationality may not cancel out and instead generate broader market patterns that differ from predictions based on hypothetical rationality.

Similarly, Becker’s argument regarding irrational consumer behavior extends to producers as well. Even firms who do not maximize profits must respond systematically to changes in their production opportunity set: As the price of inputs or the competitive conditions in the market change, even firms acting randomly, or those guided by inertia, respond accordingly. For instance, a competitive market that becomes monopolized will tend to lower output even when firms are irrational (Becker 1962). This observation is of limited antitrust significance, because it only states that markets with irrational firms generally move in the *direction* predicted by traditional models. However, antitrust law treats differently market behaviors with the same propensity—such as increase in price or a reduction in output—depending on the *magnitude* of change. In the United States, for instance, mergers among competitors are legal unless they are likely substantially to lessen competition; monopolization and attempted monopolization both apply only to firms above a certain market power threshold; and exclusive dealing, tying arrangements, and some other restraints of trade similarly are prohibited for some firms yet permitted for others depending, inter alia, on their

degree of market power. Similarly, competition law in the European Union prohibits only firms with sufficiently large market share from conduct that amounts to an abuse of their dominant position. In each case, therefore, markets that move in the same general direction will generate different legal results depending on the respective magnitude of change in market power. This change, however, may partly depend on the degree and nature of producers' rationality in a given market setting.

Even when evaluating market-wide outcomes, moreover, antitrust law ultimately is concerned with the conduct of specific firms. Yet such conduct and its competitive effects also depend on the degree to which the specific firm and other market participants adhere to the precepts of rational profit maximization. To illustrate, the same allegedly predatory conduct that could not harm competition in a world populated only by perfectly rational firms—say, because market conditions make recoupment of the costs invested in predation unlikely—can generate significant competitive harm where a real monopolist may be irrationally aggressive towards new entrants.

### 3.2.2 *Selection Mechanisms*

Competition among producers and the arbitrage activities of sophisticated actors help markets select for rationality. The competitive process may align producer behavior with rationality-based models by weeding out less capable and thus less profitable competitors who ultimately will not survive (Alchian 1950). Antitrust commentators frequently assume, in fact, that competition leads boundedly rational decision-makers to deplete their resources by making inefficient decisions while their rational competitors enjoy consistently higher profits (Bailey 2010).

Yet this selection argument is of limited significance for antitrust analysis. For one, while competition may weed out those who consistently underperform, deviations from rationality are variable and heterogeneous (Mitchell 2003; Rachlinski 2000; Tor 2008). When decision-makers exhibit different biases to different degrees at different times, even many of those who ultimately outperform their competitors may still differ substantially from the hypothetical rational actor. Even more significantly, competitive discipline penalizes only boundedly rational behaviors that reduce profitability, while promoting those that benefit market participants. For example, competitive selection punishes most some biased decision-makers who take risks that their rational competitors avoid, yet rewards some fraction of the former with higher returns, so that these particular boundedly rational competitors outperform their rational peers (Tor 2002). In addition, antitrust focuses on less competitive markets that inevitably exert more limited disciplinary pressure on market participants. A monopolist in a market with significant entry barriers that limit the efficacy of competitive discipline, for instance, can dissipate some of its supracompetitive profits monopoly by operating less efficiently (Demsetz 1982; Leibenstein 1980), whether due to systematic deviations from rationality or for other reasons.

Arbitrage by rational actors who identify, exploit, and consequently erode the profit opportunities generated by the errors of boundedly rational decision-makers may also

help select for producer rationality. But effective arbitrage necessitates a sufficiently large group of market participants who can identify the opportunity and bear the risks and costs involved with selling to or buying from their boundedly rational counterparts. It also requires the availability of substitutes for the products that boundedly rational actors over- or underprice. These conditions rarely exist even in sophisticated financial markets, however, not to mention in real product markets (Shleifer 2000).

### 3.2.3 *Firm-Level Mechanisms*

The rationality of specific market participants—as distinct from those market-level outcomes generated by aggregation and selection mechanisms—also can be material for antitrust analysis. Before turning in the next section to study those intrafirm mechanisms of rationality, the following paragraphs consider two related mechanisms that impact the firm overall, as a single unit of production in the market.

First, producers in markets primarily aim at earning profits and for this reason alone should be more rational as they try to avoid errors that plague decision-makers outside markets. Indeed, the notion that monetary incentives matter and that larger incentives lead to greater effort and better performance is near axiomatic in traditional economics. Yet the empirical evidence suggests this is not always the case: At least where individual decision-makers are concerned, financial incentives can increase effort but generate only limited improvements in intuitive judgment and decision behavior, at times even diminishing performance (Camerer and Hogarth 1999; Gneezy and Rustichini 2000). Producers' increased competitive efforts also may be directed at goals other than pure profit maximization, such as obtaining a larger share of or relative position in the market (Armstrong and Huck 2010).

Second and related, boundedly rational producers must learn to correct their mistakes if they are to improve their performance. Effective learning requires firms to identify judgment and decision errors, to associate these errors with specific negative consequences, and finally to replace them with more rational behaviors (Tversky and Kahneman 1986). However, in typical antitrust settings, such learning can be exceedingly difficult. Most judgments and decisions in product markets are made under uncertainty; outcomes are multiply-determined and delayed; feedback is limited and noisy; and there is no reliable information about the counterfactual outcomes that would have occurred had a different choice been made. Over time and with experience producers nevertheless can improve their performance even without actual learning. They may imitate successful competitors, follow established industry norms, or seek the advice of consultants with expertise in improving business outcomes. But such efforts may not always align the producers' conduct with strict rationality. Imitation may be directed at the wrong aspects of competitors' conduct, industry norms may not be rational or efficient, and to seek and to invest resources in procuring outside advice—not to mention follow it successfully—firms must first recognize their sub-optimal conduct. Finally, if the challenges involved in learning from experience in product markets were not enough, many of the most significant judgments and decisions from an antitrust perspective are infrequent, sometimes unique. Entry into new

markets, mergers and acquisitions, the development of new business strategies and vertical arrangements, and similarly uncommon decisions all offer producers only limited learning opportunities.

The various mechanisms of market rationality thus constrain some deviations from strict rationality, partly confirming the common intuition that producers are more likely than consumers to behave rationally. At the same time, the rationality-promoting effects of aggregation, selection, profit-seeking, and learning are more limited than legal analysts typically recognize, particularly in those antitrust-relevant market settings.

## 4 THE FIRM

---

In addition to the effects of markets, the judgment and decision behaviors of firms and their managers are shaped by intrafirm institutional mechanisms as well. Indeed, much like the various economic advantages of firms over individual contracting in the market that the economic theory of the firm highlights (Alchian and Demsetz 1972; Coase 1937; Jensen and Meckling 1976; Williamson 1971), one might expect firms also to exhibit superior rationality compared to individuals. After all, most firms that antitrust law is concerned with are large business organizations that can recruit expert agents to manage them; draw on organizational routines based on extensive experience to guide the behavior of managers and, even more so, to direct the conduct of lower-level employees within the organization; use sophisticated governance arrangements to align managers' judgments and decisions with the interests of the firm; and use corporate boards—groups of experienced directors who can direct, monitor, and discipline managers.

Nonetheless, the empirical evidence on managerial and firm behavior—both generally and with respect to important antitrust-relevant tasks in particular—reveals a more complex picture. While managers are sophisticated and experienced professionals, they still are human. As amply illustrated by the growing literature on behavioral corporate finance, these corporate decision-makers are selected and shaped by institutional forces to manifest greater rationality in some respects but systematic bias in others. In the same vein, the corporate governance literature demonstrates the fundamental limits of the main intrafirm governance mechanisms that may promote rationality, from contractual arrangements to corporate boards.

### 4.1 Managers in the Firm

Business managers may be more rational in their judgment and decision behavior than other individuals because of their expertise (Engel 2010a). Research shows that in some fields experts outperform individuals who do not have domain-specific expertise. But

as explained below the evidence also reveals that experts often make mistakes that resemble those of other individuals where the rationality of judgment and decision behavior specifically is concerned.

#### 4.1.1 *Managerial Expertise*

The main factors that determine experts' performance—besides the extent of their experience and their subject-matter expertise—are the nature of the task and the decision environment. The learning processes that help experts develop more rational behavior resemble those that individuals employ for learning more generally. When feedback is clear and readily available—as in weather forecasting—experts can perform well even in the face of an uncertain environment. But in many other domains—particularly where feedback is limited and ambiguous—experts frequently do not exhibit more rational behavior than nonexperts. Thus, experts require high-validity environments, in which cues are informative, and an adequate opportunity to learn if they are to act more rationally and develop “skilled intuition” (Kahneman and Klein 2009).

Unsurprisingly, therefore, numerous studies reveal systematic judgment and decision errors by experts, even while these more experienced, sophisticated actors outperform nonexperts in some settings. In fact, a leading researcher in the field recently noted that some of his earliest collaborative studies of biases that result from people's reliance on intuitive, heuristic-based judgment used experts in statistics as participants (Kahneman 2011). Other evidence shows systematic biases in the clinical judgments made by physicians and various other health professionals (Arkes et al. 1981; Chapman and Chapman 1969; McNeil et al. 1982). Furthermore, studies reveal a variety of systematic judgment and decision errors by business and finance professionals, from veteran accountants and real estate brokers, to investment managers, options traders, and financial planners (Fox, Rogers, and Tversky 1996; Joyce and Biddle 1981; Olsen 1997; Roszkowski and Snelbecker 1990). These latter findings indeed are unsurprising, given the limited and ambiguous nature of the feedback that business decision-makers frequently face.

#### 4.1.2 *Intra-firm Selection*

One might hope for a further alignment with rational models on the part of corporate managers due to selection effects. The managers whose behavior is most relevant for antitrust purposes are not only business experts; they belong to a smaller, more select group that reaches top posts on the corporate ladder presumably based on their superior performance, and may thus also be more rational. Theory and evidence both suggest, however, that managerial selection is of limited efficacy here, in part echoing the limits of market-level selection discussed above. Yet some mechanisms of rationality are even more constrained or altogether irrelevant where managers are concerned. After all, managerial behavior is a matter of individual rather than aggregate rationality. And managerial rationality is an even less likely target for successful arbitrage than firm-level conduct in product markets.



Moreover, managerial competitions at best reward those performance elements that most closely correlate with the firm's long-run profitability. Because rationality only partly correlates with the firm's success, effective competitive selection within the firm will promote managers who exhibit some systematic deviations from rationality. For instance, successful managers may benefit from a reputation for consistency and commitment, which can lead them to take into account sunk costs that rational actors would disregard (Staw 1997).

Similarly, managerial tournaments may promote overconfidence—a term denoting a cluster of loosely related deviations from rational judgment (Glaser and Weber 2010). For example, overconfident managers may persevere in difficult situations, exhibit greater ambition and confidence in their performance, and disproportionately attribute their successes to their own prowess over luck. These characteristics, however, may make such managers more attractive to the firm than their unbiased peers and thus more likely to be selected for top positions. Of course, overconfidence is not always beneficial for managers or firms. Some behavioral corporate finance studies show, for example, that banks with overconfident CEOs take greater risks than their peer institutions (Niu 2010), and top-performing mutual fund managers tend to trade more following success and exhibit worse performance when they do so (Puetz and Ruenzi 2011). Studies further show that managerial overconfidence distorts both investment and financing decisions at the firm level (Malmendier and Tate 2005); helps explain the volume, type, and financing of mergers and acquisitions activity (Ferris, Jayaraman, and Sabherwal, forthcoming); and is even linked to aggressive accounting and an increased likelihood of financial misreporting (Ahmed and Duellman 2012; Schrand and Zechman 2012).

Thus both theory and the rapidly accumulating evidence show that behavioral phenomena like managerial overconfidence exert significant, measurable effects on firm-level conduct in the market. Apparently, intrafirm competition does not eliminate all deviations from strict rationality and can even promote some of them.

## **4.2 Intrafirm Institutions**

### *4.2.1 Organizational Repairs*

Managers also may be more rational specifically because they operate within firms. First, when firms have the time and means to learn from experience and repeated feedback, they can develop “organizational repairs”—that is, internal procedures and rules that aim to overcome systematic individual shortcomings. The management literature provides anecdotal illustrations, for example, of organizations using maxims intended to remind employees not to make biased attributions, utilizing strategies aimed at collecting sufficient, relevant information, and developing methods for evaluating their information and hypotheses more objectively (Heath, Larrick, and Klayman 1998).

Nevertheless, organizational repairs have limited success and largely are unpredictable, tending to be most efficacious when based on bottom-up learning within the firm in a specific domain. These characteristics may resemble some of the tasks required of lower or midlevel employees in many firms. They do not apply, on the other hand, to many significant antitrust-relevant tasks that managers face, such as when they need to make judgments and decisions regarding the firm's overall pricing strategy, alliances with actual or potential competitors, mergers and acquisitions, and more. Such judgments and decisions are made infrequently, by top management, and usually offer limited, noisy feedback, all of which characteristics make organizational repairs unlikely.

#### 4.2.2 *Agency*

Additionally, managers may approximate rational action simply because they function as the firm's agents. There is some evidence that agents who operate on behalf of others act more rationally than individuals acting on their own behalf. The endowment effect—wherein individuals value entitlements they possess more highly than identical ones they do not hold (chapter 12 by Korobkin in this volume)—was not manifested by experimental participants taking the role of agents and transacting on behalf of their principals (Arlen, Spitzer, and Talley 2002). Behavioral evidence also suggests that egocentric biases are less likely to impact judgments made on behalf of others, at least insofar as the agent has not adopted the principal's perspective (Tor 2002).

The better alignment of agents' judgment and choice with rationality, however, would be of limited assistance to managers in overcoming those judgment and decision errors they still manifest with respect to major corporate decisions. For one, the limited evidence of agents' increased rationality still pertains to but a few of the relevant behavioral phenomena. In addition, agents' rationality advantage does not apply to most antitrust-relevant managerial tasks. The experimental elimination of agents' endowment effect, for example, was driven by participants' framing of the entitlements they traded based on the exchange value of these entitlements (Kahneman, Knetsch, and Thaler 1990; Koszegi and Rabin 2006). The impact of loss aversion on key antitrust-relevant decisions, on the other hand, concerns the managers' own strategic decisions about the overall course of the firm rather than to entitlements such as goods held by the firm for routine transactions. Agents' advantage regarding egocentric biases similarly is unlikely to pertain to judgments of their own managerial ability and expertise. Similar limitations apply more generally, whenever managers' judgments and decisions regarding their own abilities, plans, and performance are involved.

#### 4.2.3 *Corporate Boards and Governance*

Corporate managers nonetheless may exhibit superior performance because they make most significant judgments and decisions in a group of top managers or the corporate

board of directors, with the benefits of multiple viewpoints, cumulative experience, and deliberation.

Despite this claim's intuitive appeal, however, the empirical evidence does not show boards (or top management groups) are reliably likely to avoid those systematic decision errors that plague individual managers. Small-groups outperform individual rationality in some cases but at other times exhibit similar or even more extreme judgmental biases and decision errors, with their ultimate performance largely dependent on case-specific variables (Hill 1982; Kerr, MacCoun, and Kramer 1996; Tindale, Kameda, and Hinsz 2003).

Beyond their limited capacity to meliorate individuals' errors, some common characteristics of group decision-making—most notably deliberation—often cause additional, group-level biases (Berg, Dickhaut, and McCabe 1995). Groups may exhibit groupthink, promoting an erroneous consensus that does not reflect the information held by individual group members. Their deliberations, instead of leading to a superior integration of group members' information and perspectives, can also cause group polarization, so that the resulting collective view of the group is more extreme than the individual members' pre-deliberation tendencies. Hence, while senior management's collective judgment and decision-making can and will sometimes manifest superior performance, there is little reason to believe they will approximate the predictions of rational models across the board.

Finally, the extensive corporate governance scholarship suggests that in reality corporate boards possess limited efficacy in disciplining and monitoring managers (Adams, Hermalin, and Weisbach 2010; Hermalin and Weisbach 2003) and often are dominated by CEOs (Bebchuk and Fried 2004). Therefore, even with the gradual shift in recent years towards increased board power at the expense of management, corporate boards are unlikely to shape many of senior management's significant, antitrust-relevant judgments and decisions, except in some extreme cases. The rationality advantages of boards, if any, will thus impact managerial behavior only to a limited degree where behavioral antitrust is concerned.

## 5 BEHAVIORAL ANTITRUST

---

Behavioral antitrust draws on evidence suggesting that real, boundedly rational market participants deviate systematically from the predictions of the rationality-based economic models that antitrust law currently relies on. In some cases, scholars further argue that these deviations warrant changes in antitrust doctrine or enforcement policy. This section draws on some of these analyses, across different areas of antitrust law, illustrating how behavioral antitrust can and should account for those manifestations of bounded rationality that markets and firms allow and sometimes even facilitate.

## 5.1 The Market and Behavioral Antitrust

### 5.1.1 *Firms Exploiting Boundedly Rational Consumers: Aftermarket Power, Bundling, and Tying*

One of the claims commonly made by scholars applying behavioral findings—whether those who support behavioral antitrust or oppose it—is that the exploitation of boundedly rational consumers by sophisticated firms may raise antitrust concerns (Ginsburg and Moore 2010; Huffman 2012; Reeves and Stucke 2011). In particular, commentators make such arguments with respect to aftermarket power and in the case of bundling and tying practices.

The behavioral aspects of aftermarket power analysis have been raised, at least implicitly, by the US Supreme Court in *Eastman Kodak Co. v. Image Technical Services, Inc.*, 504 U.S. 451 (1992), which affirmed the denial of summary judgment to the defendant, a manufacturer of business copiers. The plaintiffs alleged that Kodak’s requirement that buyers of its machines service the copiers exclusively with the manufacturer’s original parts amounted to illegal tying under Section 1 and monopolization under Section 2 of the Sherman Act.

The *Kodak* dissent argued for summary judgment based on the assumption that consumers of copiers are rational decision-makers. In the face of rational consumers, a competitive market in copiers necessarily prevents Kodak from exercising power in the sale of parts for its own machines, even if consumers who already possess Kodak machines are “locked in” to using these parts. The dissent reasoned that any increase in the price of parts and services in the aftermarket that sought to exploit the power resulting from lock-in would effectively make Kodak’s machines more costly in the competitive primary market for copiers. Rational consumers, who take into account the future costs of parts and services over the copier’s lifetime, would find the machines less attractive. Yet if this were the case, Kodak could not charge higher prices for parts, because that would entail losing copier sales to competitors in the primary market.

This rationality-based argument was rejected, however, by the Court’s majority that ruled Kodak could have exercised power in the aftermarket for the sale of machines parts even while the primary market for copiers was competitive. The majority’s position can be explained on behavioral grounds. Though perfectly rational consumers in the primary market would have sufficed to deter Kodak from exploiting its aftermarket power, the same does not necessarily hold for boundedly rational consumers who do not incorporate all the future costs of parts over the copiers’ lifetime into the primary market price. Importantly, the majority’s conclusion did not require a positive finding that Kodak in fact exercised power in the parts aftermarket, since the Court only affirmed the denial of summary judgment by the court of appeals. Kodak’s actual aftermarket power depended on the proportion of those myopic consumers who do not take future costs effectively into account (Bar-Gill 2012a) versus their sophisticated counterparts (who do account for future costs), as well as on the intensity of

competition from other copier manufacturers for primary market sales (Bennett et al. 2010).

Thus, even without fully examining the efficiency consequences of the tying of copier parts, *Kodak* clearly illustrates how markets can encourage firms to adopt different competitive strategies depending inter alia on the degree of bounded rationality manifested by their consumers. In such cases, antitrust cannot assume that primary market competition always will prevent the exploitation of aftermarket power, nor that the market inevitably will facilitate such exploitation. Instead, the plausibility of this conduct and its competitive consequences will depend on the circumstances of the specific market at hand.

Beyond aftermarket power, antitrust scholars argued that the vertical practices of product bundling and tying may exert more powerful effects on real consumers than traditional models acknowledge. For example, behavioral findings on consumer inertia, the endowment effect, and the impact of defaults on choice indicate that consumers may find it difficult to switch from one product to a competing one, even where the objective costs of switching are small. Indeed, both theoretical arguments and experimental tests suggest that rebate schemes and other loyalty programs have stronger effects on the behavior of real consumers than rationality-based models expect them to have (Beckenkamp and Maier-Rigaud 2006). Some analysts therefore argue that dominant firms can use bundling, tying, and similar devices to foreclose competition more effectively than antitrust currently assumes (Bennett et al. 2010).

However, a more careful analysis of market dynamics suggests that the potential susceptibility of consumers to behavioral manipulation by firms will not always advantage monopolists or dominant firms. The stickiness of consumer behavior frequently redounds to incumbents' benefit because new entrants and smaller competitors may find it more difficult than standard models predict to attract consumers on the basis of lower price or higher quality alone. Yet sufficiently large multiproduct firms with a small share in a given product market may profitably expend resources on shaping consumer behavior and, consequently, exert greater competitive pressure on incumbents than commentators tend to assume when faced with boundedly rational consumers.

The example of bundling and tying reveals, therefore, that the processes through which markets facilitate the exploitation of boundedly rational consumers by firms may sometimes benefit nondominant as well as dominant firms. In these circumstances, some vertical practices that would have been unattractive to sellers in markets populated by rational consumers may turn out to be advantageous in real market settings.

### 5.1.2 *Market Discipline with Boundedly Rational Entry*

The competition among new entrants into markets illustrates how even when competition disciplines boundedly rational market participants, markets can generate outcomes that are very different from those predicted by rationality-based models. Consequently, the real social benefits and costs of entry and, importantly, its impact on incumbents' market power—a matter of importance for various areas of antitrust

law—also differ from those that antitrust traditionally assumes. Potential entry plays an important role in merger assessments, for instance, because it can counteract the anticompetitive effects of increased market power, which might otherwise follow a horizontal merger (Horizontal Merger Guidelines 2010). More generally, with effective entry even firms with large market shares may not be able to exert their market power, an essential element of the antitrust violations of monopolization and attempted monopolization, tying, exclusive dealing, and other antitrust violations (Areeda and Hovenkamp 2006; Landes and Posner 1981).

Traditional antitrust models assume that entry will only take place when it maximizes entrants' profits, requiring it to have a positive risk-adjusted, net present value (Tor 2002). The empirical evidence on entry paints a very different picture, however, showing abundant excess entry that appears unjustified based on entrants' objective prospects of survival and profitability (Baldwin 1995). Studies further reveal two additional puzzling entry phenomena: For one, entry appears insensitive to some of the economic variables that predict future profitability, from the intensity of competition to certain entry barriers (Evans and Siegfried 1992; Geroski 1991). Startup entrants, moreover, not only fail more frequently, but do so to such an extent that they obtain lower expected payoffs and thus exhibit inferior average performance compared to diversifying entrants (Dunne, Roberts, and Samuelson 1988, Dunne and Roberts 1989).

A behavioral analysis of entrants' judgments reveals, however, that the psychology of optimistic overconfidence can explain these three puzzling phenomena (Tor 2002). New entrants make their personally significant judgments of entry's prospects under conditions of extreme uncertainty so that a host of psychological processes leads them collectively to overestimate their prospects. These behavioral phenomena also reduce entrants' sensitivity to market predictors of success that should encourage or discourage entry, including the expected intensity of competition and certain entry barriers. Some of these psychological factors, moreover, exert a differential impact that makes startup entrants more biased when judging their entry prospects compared to diversifying, experienced firms that already are successful in adjacent markets (Tor 2002).

The forces that shape entrants' judgments generate a competitive landscape that differs significantly from that envisioned by traditional antitrust models. Because all entry is not the same, the more biased and numerous startups fail at greater proportions than their diversifying competitors, yet are overrepresented among those few entrants who ultimately survive and prosper. Hence the same competitive market forces that serve to punish most competitors who overestimate their prospects can still fail to prevent biased entry more generally and may even attract it by producing those uncommon instances of exceptionally successful entrants who "made it big" against the odds. At the same time, insofar as new entry is associated with innovation, particularly for startups, the biases of entrants can also generate significant benefits to society writ large.

Where the longer-term impact of entry on incumbents is concerned, the behavioral analysis of entry suggests that even while entry often is not exceptionally difficult, postentry success and survival are unlikely for the typical entrant. Most startups, and

small entrants more generally, pose no short-term competitive threat to incumbents, but some large diversifying entrants possibly do. In the long run, however, the few successful boundedly rational innovative entrants are an important source of competitive pressure on incumbents. These outcomes have important implications for antitrust law and policy. For one, they support the law's hostility to unnecessary restrictions on new entry, given its important procompetitive benefits. The law nonetheless should be wary of relying on findings of low entry barriers alone to guarantee competitive pressure on incumbents in the short run, since most new entrants detract little from incumbents' market power except where those less-common diversifying entrants are concerned.

### 5.1.3 *The Maintenance of Market Power: Bounded Rationality and Predation*

This author and others identified circumstances where monopolists may engage in predatory behavior that fails to maximize profits and is therefore irrational according to the standard account (Gerla 1985; Tor 2003). For example, managers of a dominant firm that is losing market share may take excessive risks due to loss aversion, while managers of established, profitable incumbents may exhibit the opposite pattern of excessive risk avoidance. Analysts further offer evidence from antitrust cases of predatory behavior that appears irrational on traditional accounts (Leslie 2010), while others show how even rational monopolists may find it beneficial to imitate the behavior of their irrational counterparts when market participants know that some market participants may engage in irrational predation (Edlin 2012). In these cases, therefore, market dynamics may encourage behavior that initially seems to deviate from rationality rather than discourage such conduct.

### 5.1.4 *Bounded Rationality and Unexploited Market Power*

Some commentators argue that traditional models, such as those used to evaluate the unilateral effects of horizontal mergers, can overstate the harm created by newly acquired, substantial market power (Bailey 2010). They aver that real firms sometimes avoid fully exploiting their market power, charging prices they deem "fair" instead of maximizing profits, thereby potentially justifying a more permissive approach to merger enforcement.

This argument may be compelling when market power is generated by recent, temporary changes in market conditions. In such cases, firms may not exercise their power fully, whether to maintain a reputation for offering low prices or to avoid negative reactions by consumers to prices the latter perceive as "unfair" (Kahneman, Knetsch, and Thaler 1986). On the other hand, the powerful pressure towards greater profitability that markets and firms exert suggests that unexploited long-term market power for fairness concerns should be uncommon. For one, markets and firms may penalize managers who fail to extract value from patently available market power. The behavioral force of adaptation also suggests that customers—even those who initially view the exercise of market power as unfair—will tend to accept it as normal over time.



Hence the same forces that militate against short-term exploitation of such power dissipate and allow, even support, the profitable exercise of market power in the longer run.

## 5.2 The Firm and Behavioral Antitrust

### 5.2.1 *Privately Beneficial Bounded Rationality: Cartels*

Where horizontal restraints among competitors are concerned, a number of scholars argue that behavioral forces increase the likelihood and stability of cartelization beyond what traditional antitrust theories suggest (Armstrong and Huck 2010; Leslie 2010). Traditional economic accounts consider cartels unlikely, for instance, where products are differentiated, competitors numerous, or entry barriers low. Cartels, moreover, should be inherently unstable, because cartel members find it most profitable to cheat on the cartel and—although they would all be better off with a cartel than without it—rational cartelists anticipate cheating by their counterparts and avoid cartelizing or abandon it quickly when cheating is hard to detect or difficult to enforce against. However, there are by now many examples of real-world cartels found in industries and product markets where they should not have existed and could not have thrived for extended periods according to traditional economic accounts (Armstrong and Huck 2010; Reeves and Stucke 2011). These ubiquitous cartels, for instance, spanned markets with large numbers of competitors, limited entry barriers, nonhomogenous products with complex pricing and cost structures, and other characteristics that make cartelization unlikely for strictly rational actors (Levenstein and Suslow 2006; Stigler 1964).

Explaining the cartel evidence, commentators argue that behavioral factors, such as managers' social preferences for trust and cooperation, personal relationships, social networks and social norms all help competing firms establish and maintain collusive arrangements where rationality-based models that ignore these behavioral factors expect them to fail (Armstrong and Huck 2010; Bennett et al. 2010). Other researchers point to additional, nonsocial phenomena, such as managers' aspiration to obtain merely satisfactory profits, which makes the potential for secure profits through cartelization more attractive than for the traditionally assumed profit-maximizing firm (Dixon 2000; Simon 1976). The influence of these various factors thus illustrates how some characteristics that make individuals more effective managers or even just more successful in intrafirm competitions can also facilitate the creation and maintenance of cartels beyond what rationality-based models predict. Of course, firms benefiting from collusive arrangements also are unlikely to combat managerial behaviors that facilitate cartelization.

Other researchers argue to the contrary, however, that behavioral forces should destabilize collusion due to processes that traditional models ignore. For example, some studies of experimental markets show that an increase in the amount of information available to competitors about rivals' output and profits—which would make

easier oligopolistic coordination by rational actors—can lead in fact to less collusive, more competitive market behavior (Huck, Normann, and Oechssler 1999, 2000; Offerman, Potters, and Sonnemans 2002). Similarly, the broader behavioral literature makes clear that individuals' concern for relative—as opposed to absolute—outcomes is ubiquitous, particularly common in competitive settings (Garcia and Tor 2007, 2009; Garcia, Tor, and Schiff 2013), and evidenced in managerial behavior. Yet the concern for relative outcomes, if manifested by managers when making decisions on behalf of their firms, can sometimes destabilize cartels and make markets more competitive (Armstrong and Huck 2010), although it is difficult to obtain empirical evidence for such patterns in real-world markets.

The seemingly conflicting findings can be reconciled, however. As noted already, the behavior of antitrust actors is variable, heterogeneous, and shaped by market and firm institutions. Hence while some markets and firms may make collusion more likely than typically assumed, others may make it more difficult and less likely. More generally, however, beyond showing how specific firm and market characteristics may facilitate or inhibit collusion in ways that traditional models neglect, the behavioral evidence also reveals that established patterns of market behavior—whether competitive or collusive—exhibit greater stability than standard antitrust models assume. Extant rivals are slower to recognize and embark on mutually profitable opportunities for coordinated behavior—whether legal, collaborative arrangements or illegal cartelization—than rationality-based accounts allow for. At the same time, established collaborative or collusive arrangements are more stable than they would have been if the parties were strictly rational actors (Aviram and Tor 2004).

A number of factors combine to make market behavior “sticky.” In the domain of judgment, established norms of rivalry diminish competitors' ability to identify profitable opportunities for cooperation and vice versa where collusive norms prevail. Managers' risk attitudes can also cause them to overestimate the value of extant arrangements and underestimate the benefits of alternative courses of interaction with rivals. Moreover, some decision phenomena—including the status quo bias (chapter 11 by Zamir in this volume) and the aversion to comparative ambiguity (Camerer and Weber 1992)—may lead competitors consciously to forgo risky, profitable opportunities for collusion while also inhibiting cartelists' willingness to chance potentially profitable competitive alternatives to ongoing collusive arrangements (Aviram and Tor 2004).

### 5.2.2 *Divergence of Private Benefits between Managers and Their Firms: Mergers*

Unlike those deviations from strict rationality that benefit both manager and firm and may therefore be allowed or even encouraged by intrafirm processes, other manifestations of bounded rationality may be costly to the firm. Firms might be expected to develop mechanisms that address such behaviors, yet the empirical evidence suggests they do not always do so, at least not effectively. When the private benefits to managers diverge from those of the firm, boundedly rational behavior may be viewed as

another aspect of the familiar agency problem between the two parties (Easterbrook and Fischel 1991), which firms and the law both seek to minimize but cannot eradicate altogether.

With respect to merger policy, commentators draw on empirical evidence from the corporate finance literature as well as on some behavioral findings to note that many mergers prove inefficient rather than profit-maximizing, as antitrust now assumes (Horton 2011; Mueller 2003; Reeves and Stucke 2011), and at least partly result from managerial overconfidence (Malmendier and Tate 2005; Roll 1986). Empirical studies found, for instance, that mergers often diminish rather than increase the market value of the acquiring firm, and behavioral research long has suggested that excess merger activity is likely driven inter alia by the optimistic overconfidence of managers (Moeller, Schlingemann, and Stultz 2005; Roll 1986). In a similar vein, scholars contend that merger-specific efficiencies—which parties proffer to show how transactions that raise competitive concerns generate benefits—are not only difficult to substantiate but often fail to materialize (Horizontal Merger Guidelines 2010; Oldale 2010; Ravenscraft and Scherer 2001).

Yet, even among those who note the prevalence of inefficient mergers, opinions diverge as to whether this systematic deviation from standard economic models matters for antitrust law. Some argue that an accounting for the overall efficiency of a proposed merger is outside antitrust's limited mandate to prevent only mergers "substantially likely to lessen competition" (Werden, Froeb, and Shor 2011). They also aver that the antitrust agencies already are skeptical regarding claims of merger-specific efficiencies (Oldale 2010; Werden, Froeb, and Shor 2011). Others counter that the evidence of prevalent inefficient mergers justifies closer scrutiny by the agencies of transactions with potentially anticompetitive effects. After all, merger policy balances the uncertain prospects of over- and underenforcement—that is, the risk of blocking efficient mergers versus the risk of allowing the consummation of anticompetitive ones. Hence, these commentators assert, if inefficient mergers are prevalent, the risks of overenforcement diminish and a greater emphasis on preventing anticompetitive mergers is warranted (Reeves and Stucke 2011).

Irrespective of its ultimate antitrust evaluation, therefore, the prevalence of inefficient mergers clearly illustrates the limited capacity of firms to eliminate costly managerial behavior, particularly where the interests of managers and their firms diverge. Indeed, in the case of mergers, at least some of those intrafirm governance mechanisms that seek to address agency problems—like the contractual arrangements that determine managerial compensation in part based on firm size—may both promote inefficient mergers by strictly rational managers and do little to discipline boundedly rational merger activity.

### 5.2.3 *Costly Bounded Rationality and the Limits of the Firm: Resale Price Maintenance*

At times, the bounded rationality of managers harms the firm without benefiting managers, suggesting that firms are more likely to discipline such deviations from strict

rationality. Besides the general limitations of intrafirm mechanisms, however, firms must first identify their managers' costly errors and determine how to remedy them, a task that can be particularly challenging under some circumstances, as exemplified by the case of resale price maintenance (RPM)—a vertical agreement in which manufacturers dictate the minimum price at which their products are resold at retail.

Antitrust commentators have long debated the merits and demerits of this vertical practice based on the assumption of manufacturer rationality. Some point to RPM's potential procompetitive benefits, primarily the stimulation of various costly retail services that enhance consumer demand and total sales of manufacturers' products (Easterbrook 1984; Klein 2009). Others note the anticompetitive potential of the practice, which can facilitate cartels and otherwise soften horizontal competition at manufacture or retail (Comanor 1985; Cooper et al. 2005). On either account, therefore, RPM always benefits its adopters, whether by beneficially raising demand or harmfully increasing market power. Because it can generate pro- and anticompetitive outcomes alike, resale price maintenance is now governed by the rule of reason, under which plaintiffs must show the practice anticompetitive on balance in the specific case (*Leegin* 2007).

Behavioral antitrust offers a novel perspective on RPM. Specifically, both the historical evidence on the practice and the behavioral literature reveal that when resale price maintenance is legal some manufacturers will excessively impose it on their retailers (Tor and Rinner 2011). Manufacturers are prone to error with respect to vertical price restraints due to a number of behavioral phenomena: Judgmental biases—including anchoring on biased information from retailers, the availability and representativeness effects of a few vivid cases and small samples—lead managers to overestimate the expected harms of retailer price-cutting. Loss aversion makes managers averse to the potential harms of price-cutting, and fairness concerns can lead them to react negatively to perceived free-riding by discounting retailers on the manufacturers' brand investments. Manufacturers also tend to overvalue the direct elimination of retail price-cutting through RPM and to overestimate its benefits compared to alternative distribution arrangements that address the potential harms of price cutting in more indirect ways. Unlike their perfectly rational, hypothetical counterparts, therefore, real manufacturers sometimes overuse RPM, often at a cost to their own sales and profitability. Excessive use of the practice, moreover, can harm some retailers who are forced to charge higher prices for the manufacturer's products than is optimal for them in their specific market setting.

These potential costs of boundedly rational RPM notwithstanding, further analysis reveals that the practice raises antitrust concerns only in those limited circumstances when it harms the competitive process itself rather than being costly to a few specific manufacturers and some of their retailers. Such harm to competition may occur, for example, when firms with substantial market power excessively use RPM or when the practice is pervasive in a given market, circumstances under which consumers may not have effective substitutes for the price-maintained products. Based on these and similar observations, moreover, behavioral antitrust can offer an appropriately structured

rule of reason for RPM, following the recent approach taken by the US Supreme Court (*Leegin* 2007).

Importantly, the behavioral analysis of resale price maintenance also illustrates how firms can be slow to discipline some behavioral phenomena. The historical data shows that when the practice is legal certain manufacturers employ it for significant periods of time, even when it is not beneficial to them. The slow learning of some firms to avoid resale price maintenance is understandable, however, once the relevant behavioral processes are taken into account. On the part of firms, at least some of the forces that make RPM excessively attractive—such as the desire to control and limit business risks—can be pervasive features of managerial behavior more generally (Kahneman and Lovallo 1993; March and Shapira 1987; Shapira 1995). Furthermore, when a business practice is unprofitable, managers may learn of it over time and avoid it. However, firms find it particularly difficult to learn the precise consequences of RPM. Manufacturers expect resale price maintenance to be costly in the short run, because the practice tends to increase retail prices and therefore reduce consumer demand. Yet the purported long-term benefits of RPM are hard to ascertain because of the myriad factors that over time impact consumer demand for a given product and the competitive conditions in the market. Moreover, the imposition of vertical price restraints itself impedes firms' ability to learn when RPM is beneficial or costly for them. This practice, by definition, eliminates retail price variations for the manufacturer's products and, with them, the valuable information they could have provided regarding the response of different markets to different price points (Tor and Rinner 2011).

## 6 CONCLUSION

---

The behavioral antitrust enterprise is still nascent. Despite the recent flood of scholarship seeking to apply behavioral findings to antitrust issues, many important questions have yet to be examined at any depth. And while the interest of scholars in this new approach promises to yield further insights, the challenges of effectively and convincingly applying to antitrust those behavioral findings that primarily concern the non-market behavior of individual decision-makers are significant. Partly for this reason, commentators recently have called for additional empirical evidence to shed light on behavioral antitrust questions (Bennett et al. 2010; Ginsburg and Moore 2010; Reeves and Stucke 2010). Indeed, some analysts have repeated the call for empirical evidence on numerous occasions without considering fully what new evidence is needed and under what conditions this evidence might help resolve the challenges facing behavioral antitrust (Reeves 2010; Stucke 2009, 2011, 2013).

For example, the suggestion that the antitrust agencies conduct additional retrospective studies of mergers in different industries—whether of mergers that were not challenged or those that were challenged unsuccessfully—would surely be beneficial (Reeves and Stucke 2011; Stucke 2009, 2011). After all, any additional evidence

regarding the real outcomes of these mergers would be informative for antitrust policy. Notwithstanding their general value for antitrust enforcement, however, this and similar types of empirical evidence will do little to clarify when and how specific behavioral phenomena shape the competitive conduct of firms in markets. Even a finding that some horizontal mergers that *ex ante* raised significant competitive concerns but ultimately went forward without enforcement action led to substantial increases in market power but no appreciable efficiencies would not contribute much to the behavioral antitrust debate. Such a finding might show the agencies are too permissive or maybe properly cautious—once the costs of erroneously excessive enforcement are taken into account—but would not prove whether this outcome reflects rational calculation (the merged firm is now better off), managerial overconfidence (efficiencies were not realized), or simply the resolution of business uncertainty that occasionally selects for less-likely outcomes.

This is not to say that amassing further empirical evidence would not advance antitrust law and policy, only that such findings will rarely clarify the extent to which behavioral phenomena manifest in antitrust markets and the precise form they take in these settings, if any. An exception to the limited behavioral utility of such findings might be where robust empirical patterns contradict standard accounts and better conform to behavioral ones. Yet even in these uncommon cases the behavioral lesson may still be ambiguous, given the uncertainty of markets on the one hand and the lack of direct evidence linking specific managerial-level phenomena with the competitive conduct and outcomes of firms on the other.

Therefore, to advance behavioral antitrust beyond a reasoned and careful theoretical application of more general findings from the lab and the field, scholars need something more than empirical antitrust evidence writ large. Experimental studies of antitrust offer one valuable and developing category of research that may yield such benefits (Engel 2005, 2007, 2009, 2010b; Engel and Zhurakhovska 2011; Hinloopen and Normann 2009). In particular, controlled experiments that test for the manifestation of more broadly known behavioral phenomena or at least are informed by these findings—such as the approach promoted recently by advocates of “behavioral industrial organization”—can offer more direct evidence that links behavioral patterns to the conduct of and outcomes for firms in markets (Van Damme, Larouche, and Müller 2009). The development of empirical studies and theoretical models that link specific managerial characteristics with firm-level outcomes—following the success of similar approaches in behavioral finance—might offer another potentially beneficial venue for future, behaviorally informed research in antitrust (Malmendier and Tate 2005; Niu 2010; Puetz and Ruenzi 2011)

Overall, different types of empirical—observational and experimental—evidence offer different benefits and limits from the perspective of behavioral antitrust. None of these various sources of evidence is likely to suffice on its own, but together they could dramatically enhance our understanding of market behavior and its competitive consequences. Even in the best of future cases, however, dramatic developments using new approaches will likely take many years to generate a sizable body of evidence. In



the interim, and probably thereafter as well, antitrust analysis will benefit greatly from a better understanding of the nature of the empirical behavioral findings that underlie behavioral antitrust.

Most importantly, this evidence concerns specific behavioral regularities (and thus also irregularities) rather than broad certainties in the mold of the universal, hypothetical model of antitrust rationality. Human behavior is highly variable and context dependent (Mitchell 2003; Rachlinski 2000; Tor 2008), so we cannot expect antitrust actors all to behave in the same fashion all the time. The institutions of antitrust—most notably the markets and firms discussed here—can make the behavior of antitrust actors more universal and predictable (Engel 2005). Yet not only are they incapable of altogether eliminating the variability and heterogeneity of individual behavior, but markets and firms also introduce some additional behavioral complexities, as this chapter made clear.

For this reason, one immediate implication of behavioral antitrust is that the law must account for both behavioral regularities and behavioral irregularities. When fashioning antitrust doctrines and enforcement policies, courts and agencies will do well to factor in the likelihood of systematic deviations from assumptions of rationality on the part of consumers and firms overall. At the same time, however, one should not mistakenly assume that these population-level tendencies apply to all market participants, all the time. Hence, those evidentiary and procedural rules, as well the decisions of courts and agencies in specific antitrust cases, should leave appropriate room for evidence of the potentially irregular conduct of specific market participants and its competitive effects.

## REFERENCES

- Adams, R. E., B. E. Hermalin, and M. S. Weisbach. 2010. The role of boards of directors in corporate governance: a conceptual framework and survey. *Journal of Economic Literature* 48:58–107.
- Ahmed, A. S., and S. Duellman. 2012. Managerial overconfidence and accounting conservatism. *Journal of Accounting Research* 51:1–30.
- Alchian, A. A. 1950. Uncertainty, evolution, and economic theory. *Journal of Political Economy* 58:211–21.
- Alchian, A. A., and H. Demsetz. 1972. Production, information costs, and economic organization. *American Economic Review* 62:777–95.
- Areeda, P. E., and H. Hovenkamp. 2006. *Antitrust law: An analysis of antitrust principles and their application*. 3rd ed. Vol. 2. New York: Aspen.
- Arkes, H. R., R. L. Wortmann, P. D. Saville, and A. R. Harkness. 1981. Hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology* 66:252–54.
- Arlen, J. 1998. Comment: The future of behavioral economic analysis of law. *Vanderbilt Law Review* 51:1765–95.
- Arlen, J., M. L. Spitzer, and E. L. Talley. 2002. Endowment effects within corporate agency relationships. *Journal of Legal Studies* 31:1–37.



- Armstrong, M., and S. Huck. 2010. Behavioral economics as applied to firms: A primer. *Competition Policy International* 6:3–45.
- Aviram, A., and A. Tor. 2004. Overcoming impediments to information sharing. *Alabama Law Review* 55:231–79.
- Ayal, A. 2011. Harmful freedom of choice: Lessons from the cellphone market. *Law and Contemporary Problems* 74:91–131.
- Bailey, E. M. June 2010. Behavioral economics: Implications for antitrust practitioners. *The Antitrust Source* 9:1–8.
- Baldwin, J. R. 1995. *The dynamics of industrial competition: A North American perspective*. Cambridge: Cambridge University Press.
- Bar-Gill, O. 2012. Competition and Consumer Protection: A Behavioral Economics Account. In *The Pros and Cons of Consumer Protection*. Swedish Competition Authority, 12–43.
- Bebchuk, L. A., and J. M. Fried. 2004. *Pay without performance: The unfulfilled promise of executive compensation*. Cambridge, MA: Harvard University Press.
- Beckenkamp, M., and F. P. Maier-Rigaud. 2006. An experimental investigation of Article 82 rebate schemes. *Competition Law Review Supplement* 2:1–29.
- Becker, G. S. 1962. Irrational behavior and economic theory. *Journal of Political Economy* 70:1–13.
- Bennett, M., J. Fingleton, A. Fletcher, L. Hurley, and D. Ruck. 2010. What does behavioral economics mean for competition policy? *Competition Policy International* 6:111–37.
- Berg, J., J. Dickhaut, and K. McCabe. 1995. The individual versus the aggregate. In *Judgment and decision research in accounting and auditing*, R. H. Ashton and A. H. Ashton. Cambridge: Cambridge University Press, 102–35.
- Brooke Group Ltd. v. Brown & Williamson Tobacco Corp.*, 509 U.S. 209 (1993).
- Camerer, C. F., and R. M. Hogarth. 1999. The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty* 19:7–42.
- Camerer, C. F., and M. Weber. 1992. Recent developments in modeling preferences: uncertainty and ambiguity. *Journal of Risk and Uncertainty* 5:325–70.
- Chapman, L. J., and J. P. Chapman. 1969. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology* 74:271–80.
- Coase, R. H. 1937. The nature of the firm. *Economica* 4:386–405.
- Comanor, W. S. 1985. Vertical price-fixing, vertical market restrictions, and the new antitrust policy. *Harvard Law Review* 98:983–1002.
- Cooper, J. C., L. M. Froeb, D. P. O'Brien, and M. G. Vita. 2005. A comparative study of United States and European Union approaches to vertical policy. *George Mason Law Review* 13:289–308.
- Demsetz, H. 1982. *Economic, legal, and political dimensions of competition*. Amsterdam: North-Holland.
- Dixon, H. D. 2000. Keeping up with the Joneses: Competition and the evolution of collusion. *Journal of Economic Behavior and Organization* 43:223–38.
- Dunne, T., and M. J. Roberts. 1989. Variation in producer turnover across U.S. manufacturing industries. In *Entry and market contestability: An international comparison*, P. A. Geroski and J. Schwalbach. Oxford: Basil Blackwell, 187–203.
- Dunne, T., M. J. Roberts, and L. Samuelson. 1988. Patterns of firm entry and exit in U.S. manufacturing industries. *RAND Journal of Economics* 19:495–515.

- Easterbrook, F. H. 1984. Vertical arrangements and the rule of reason. *Antitrust Law Journal* 53:135–74.
- Easterbrook, F. H., and D. R. Fischel. 1991. *The economic structure of corporate law*. Cambridge, MA: Harvard University Press. *Eastman Kodak Co. v. Image Technical Services, Inc.*, 504 U.S. 451 (1992).
- Edlin, A. 2012. Predatory pricing. In *Research handbook on the economics of antitrust law*, E. R. Elhage. Northampton: Edward Elgar, 144–73.
- Engel, C. 2005. *Generating predictability: Institutional analysis and design*. Cambridge: Cambridge University Press.
- Engel, C. 2007. Tacit collusion: The neglected experimental evidence. Preprints of the Max Planck Institute for Research on Collective Goods, Bonn, 2007/14.
- Engel, C. 2009. Competition as a socially desirable dilemma: Theory vs. experimental evidence. In *Competition policy and the economic approach: Foundations and limitations*, J. Drexler, W. Kerber, and R. Podszun. Northampton: Edward Elgar, 245–69.
- Engel, C. 2010a. The behaviour of corporate actors: How much can we learn from the experimental literature?. *Journal of Institutional Economics* 6:445–75.
- Engel, C. 2010b. An experimental contribution to the revision of the guidelines on research and development agreements. Preprints of the Max Planck Institute for Research on Collective Goods, Bonn, 2010/48.
- Engel, C., and L. Zhurakhovska. 2011. Oligopoly as a socially embedded dilemma: An experiment. Preprints of the Max Planck Institute for Research on Collective Goods, Bonn, 2011/1.
- Evans, L. B., and J. J. Siegfried. 1992. Entry and exit in United States manufacturing industries from 1997 to 1982. In *Empirical studies in industrial organization: Essays in honor of Leonard W. Weiss*, D. B. Audretsch and J. J. Siegfried. Dordrecht: Kluwer Academic Publishers, 253–73.
- Ferris, S. P., N. Jayaraman, and S. Sabherwal. Forthcoming. CEO overconfidence and international merger and acquisition activity. *Journal of Financial and Quantitative Analysis* 48:137–64.
- Fox, C. R., B. A. Rogers, and A. Tversky. 1996. Options traders exhibit subadditive decision weights. *Journal of Risk and Uncertainty* 13:5–17.
- Friedman, M. 1953. The methodology of positive economics. In *Essays in positive economics*. Chicago: University of Chicago Press, 3–43.
- Gabaix, X., and D. Laibson. 2006. Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics* 121:505–40.
- Garcia, S. M., and A. Tor. 2007. Rankings, standards, and competition: Task vs. scale comparisons. *Organizational Behavior and Human Decision Processes* 102:95–108.
- Garcia, S. M., and A. Tor. 2009. The n-effect: More competitors, less competition. *Psychological Science* 20:871–77.
- Garcia, S. M., A. Tor, and T. M. Schiff. 2013. The Psychology of Competition: A Social Comparison Perspective. *Perspectives on Psychological Science* 8:634–650.
- Gerla, H. S. 1985. The psychology of predatory pricing: Why predatory pricing pays. *Southwestern Law Journal* 39:755–80.
- Geroski, P. A. 1991. Some data-driven reflections on the entry process. In *Entry and market contestability: An international comparison*, P. A. Geroski and J. Schwalbach. Oxford: Basil Blackwell, 282–96.

- Ginsburg, D. H., and D. W. Moore. 2010. The future of behavioral economics in antitrust jurisprudence. *Competition Policy International* 6:89–102.
- Glaser, M., and M. Weber. 2010. Overconfidence. In *Behavioral finance: Investors, corporations, and markets*, H. K. Baker and J. Nofsinger. Hoboken, NJ: John Wiley & Sons, 241–58.
- Gneezy, U., and A. Rustichini. 2000. Pay enough or don't pay at all. *Quarterly Journal of Economics* 115:791–810.
- Heath, C., R. P. Larrick, and J. Klayman. 1998. Cognitive repairs: How organizational practices can compensate for individual shortcomings. *Research in Organizational Behavior* 20:1–38.
- Hermalin, B. E., and M. S. Weisbach. 2003. Boards of directors as an endogenously determined institution: A survey of the economic literature. *Federal Reserve Bank of New York Economic Policy Review* 9:7–26.
- Hill, G. W. 1982. Group versus individual performance: Are  $n + 1$  heads better than one?. *Psychological Bulletin* 91:517–39.
- Hinloopen, J., and H-T. Normann. 2009. *Experiments and competition policy*. New York: Cambridge University Press.
- Horton, T. J. 2011. The coming extinction of homo economicus and the eclipse of the Chicago school of antitrust: Applying evolutionary biology to structural and behavioral antitrust analyses. *Loyola University Chicago Law Journal* 42:469–522.
- Huck, S., H-T. Normann, and J. Oechssler. 1999. Learning in Cournot oligopoly—An experiment. *Economic Journal* 109:C80–C95.
- Huck, S., H-T. Normann, and J. Oechssler. 2000. Does information about competitors' actions increase or decrease competition in experimental oligopoly markets? *International Journal of Industrial Organization* 18:39–57.
- Huffman, M. 2012. Marrying neo-Chicago with behavioral antitrust. *Antitrust Law Journal* 78:105–45.
- Jensen, M. C., and W. H. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3:305–60.
- Jolls, C., C. R. Sunstein, and R. Thaler. 1998. Behavioral approach to law and economics. *Stanford Law Review* 50:1471–550.
- Joyce, E. J., and G. C. Biddle. 1981. Anchoring and adjustment in probabilistic inference in auditing. *Journal of Accounting Research* 19:120–45.
- Kahneman, D. 2011. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D., and G. Klein. 2009. Conditions for intuitive expertise: A failure to disagree. *American Psychologist* 64:515–26.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1986. Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review* 76:728–41.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1990. Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy* 98:1325–48.
- Kahneman, D., and D. Lovallo. 1993. Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science* 39:17–31.
- Kerr, N. L., R. J. MacCoun, and G. P. Kramer. 1996. Bias in judgment: comparing individuals and groups. *Psychological Review* 103:687–719.
- Klein, B. 2009. Competitive resale price maintenance in the absence of free riding. *Antitrust Law Journal* 76:431–81.

- Koszegi, B., and M. Rabin. 2006. A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics* 121:1133–65.
- Landes, W. M., and R. A. Posner. 1981. Market power in antitrust cases. *Harvard Law Review* 94:937–96.
- Langevoort, D. C. 1998. Behavioral theories of judgment and decision making in legal scholarship: A literature review. *Vanderbilt Law Review* 51:1499–540.
- Leegin Creative Leather Products, Inc. v. PSKS, Inc.*, 551 U.S. 877 (2007).
- Leibenstein, H. 1980. *Beyond economic man: A new foundation for economics*. Cambridge, MA: Harvard University Press.
- Leslie, C. R. 2010. Rationality analysis in antitrust. *University of Pennsylvania Law Review* 158:261–353.
- Levenstein, M. C., and V. Y. Suslow. 2006. What determines cartel success? *Journal of Economic Literature* 44:43–95.
- Malmendier, U., and G. Tate. 2005. Does overconfidence affect corporate investment? CEO overconfidence measures revisited. *European Financial Management* 11:649–59.
- March, J. G., and Z. Shapira. 1987. Managerial perspectives on risk and risk taking. *Management Science* 33:1404–18.
- Matsushita Electric Industrial Co. v. Zenith Radio Corp.*, 475 U.S. 574 (1986).
- McNeil, B. J., S. G. Pauker, H. C. Sox Jr., and A. Tversky. 1982. On the elicitation of preferences for alternative therapies. *New England Journal of Medicine* 306:1259–62.
- Mitchell, G. 2002. Taking behavioralism too seriously? The unwarranted pessimism of the new behavioral analysis of law. *William and Mary Law Review* 43:1907–2021.
- Mitchell, G. 2003. Why law and economics' perfect rationality should not be traded for behavioral law and economics' equal incompetence. *Georgetown Law Journal* 91:67–167.
- Moeller, S. B., F. P. Schlingemann, and R. M. Stultz. 2005. Wealth destruction on a massive scale? A study of acquiring-firm returns in the recent merger wave. *Journal of Finance* 60:757–82.
- Mueller, D. C. 2003. The finance literature on mergers: A critical survey. In *Competition, monopoly and corporate governance: Essays in honour of Keith Cowling*, M. Waterson. Cheltenham: Edward Elgar, 161–205.
- Niu, J. 2010. The effect of CEO overconfidence on bank risk taking. *Economics Bulletin* 30:3288–99.
- Offerman, T., J. Potters, and J. Sonnemans. 2002. Imitation and belief learning in an oligopoly experiment. *Review of Economic Studies* 69:973–97.
- Oldale, A. 2010. Behavioral economics and merger analysis. *Competition Policy International* 6:139–43.
- Olsen, R. A. 1997. Desirability bias among professional investment managers: Some evidence from experts. *Journal of Behavioral Decision Making* 10:65–72.
- Puetz, A., and S. Ruenzi. 2011. Overconfidence among professional investors: Evidence from mutual fund managers. *Journal of Business Finance and Accounting* 38:684–712.
- Rachlinski, J. J. 2000. The “new” law and psychology: A reply to critics, skeptics, and cautious supporters. *Cornell Law Review* 85:739–66.
- Ravenscraft, D. J., and F. M. Scherer. 2001. Some principles for post-Chicago antitrust analysis. *Case Western Law Review* 52:5–23.
- Reeves, A. P. 2010. Behavioral antitrust: Unanswered questions on the horizon. *Antitrust Source*, June.

- Reeves, A. P., and M. E. Stucke. 2011. Behavioral antitrust. *Indiana Law Journal* 86:1527–86.
- Roll, R. 1986. The hubris hypothesis of corporate takeovers. *Journal of Business* 59:197–216.
- Roszkowski, M. J., and G. E. Snelbecker. 1990. Effects of “framing” on measures of risk tolerance: Financial planners are not immune. *Journal of Behavioral Economics* 19: 237–46.
- Schrand, C. M., and S. L. Zechman. 2012. Executive overconfidence and the slippery slope to financial misreporting. *Journal of Accounting and Economics* 53:311–29.
- Shapira, Z. 1995. *Risk taking: A managerial perspective*. New York: Russell Sage Foundation.
- Shleifer, A. 2000. *Inefficient markets: An introduction to behavioral finance*. New York: Oxford University Press.
- Simon, H. A. 1976. *Administrative behavior: A study of decision-making processes in administrative organization*. 3rd. ed. New York: Free Press.
- Staw, B. M. 1997. The escalation of commitment: An update and appraisal. In *Organizational decision making*, Z. Shapira. Cambridge: Cambridge University Press, 191–215.
- Stigler, G. J. 1961. The economics of information. *Journal of Political Economy* 69:213–25.
- Stigler, G. J. 1964. A theory of oligopoly. *Journal of Political Economy* 72:44–61.
- Stucke, M. E. 2009. New antitrust realism. *Global Competition Policy* Jan. 2009:2.
- Stucke, M. E. 2010. Money, is that what I want? Competition policy and the role of behavioral economics. *Santa Clara Law Review* 50:893–979.
- Stucke, M. E. 2011. Reconsidering competition. *Mississippi Law Journal* 81:107–88.
- Stucke, M. E. 2013. Greater international convergence and the behavioral antitrust gambit. In *International Research Handbook on Competition Law*, A. Ezrachi. Northampton: Edward Elgar.
- Tindale, R. S., T. Kameda, and V. B. Hinsz. 2003. Group decision making. In *The Sage handbook of social psychology*, M. A. Hogg and J. Cooper. London: Sage Publications, 381–403.
- Tor, A. 2002. The fable of entry: Bounded rationality, market discipline, and legal policy. *Michigan Law Review* 101:482–568.
- Tor, A. 2003. Illustrating a behaviorally informed approach to antitrust law: The case of predatory pricing. *Antitrust* 18:52–59.
- Tor, A. 2004. A behavioral approach to antitrust law and economics. *Consumer Policy Review* 14:18–21.
- Tor, A. 2008. The methodology of the behavioral analysis of law. *Haifa Law Review* 12:237–327.
- Tor, A. 2013. Manuscript under review. Understanding behavioral antitrust.
- Tor, A., and W. J. Rinner. 2011. Behavioral antitrust: A new approach to the rule of reason after Leegin. *University of Illinois Law Review* 2011:805–64.
- Tversky, A., and D. Kahneman. 1986. Rational choice and the framing of decisions. In *Rational choice: The contrast between economics and psychology*. Chicago: University of Chicago Press, 67–94.
- US Department of Justice and Federal Trade Commission. 2010. *Horizontal merger guidelines*. Available at <http://www.justice.gov/atr/public/guidelines/hmg-2010.pdf>.
- Van Damme, E., P. Larouche, and W. Müller. 2009. Abuse of a dominant position: Cases and experiments. In *Experiments and competition policy*, J. Hinloopen and H. T. Normann. New York: Cambridge University Press, 107–59.
- Werden, G. J., and L. M. Froeb. 2008. Unilateral competitive effects of horizontal mergers. In *Handbook of Antitrust Economics*, P. Buccirossi. Cambridge, MA: MIT Press, 43–104.

- Werden, G. J., L. M. Froeb, and M. Shor. 2011. Behavioral antitrust and merger control. *Journal of Institutional and Theoretical Economics* 167:126–42.
- Weyerhaeuser Co. v. Ross-Simmons Hardwood Lumber Co.*, 549 U.S. 312 (2007).
- Williamson, O. E. 1971. The vertical integration of production: Market failure considerations. *American Economic Review* 61:112–23.
- Wright J. D., and J. E. Stone II. 2012. Misbehavioral economics: The case against behavioral antitrust. *Cardozo Law Review* 33:1517–53.

## CHAPTER 22

---

# BEHAVIORAL ANALYSIS OF CRIMINAL LAW

## *A Survey*

---

ALON HAREL

### 1 INTRODUCTION

---

BEHAVIORAL analysis of criminal law exploits social science methodologies (behavioral economics, psychology, and even sociology) to explore the effects of criminal law norms on criminals, judges, juries, and other decision-makers, to determine the optimal type and size of criminal sanctions, and to identify the optimal design of the enforcement system and the rules of evidence. Behavioral analysis of criminal law often addresses, criticizes, or complements the findings of the traditional economic tools by using social sciences findings concerning the content of individuals' beliefs, and the content of their preferences. As criminals, victims of crime, judges, and other relevant agents form beliefs concerning probability of detection and conviction, and those beliefs affect the propensity to commit crimes, enforcement policy, evidence law, and procedural law are as relevant to the understanding of the effects of criminal law as the substantive doctrines of criminal law itself. Hence both the economic and the behavioral approaches to the analysis of criminal law challenge the traditional doctrinal distinctions between criminal law, criminal procedure and evidence, and, last, the enforcement policy.

The behavioral approach to criminal law is founded on the research of behavioral economists, psychologists, and sociologists.<sup>1</sup> Unlike traditional neoclassical

<sup>1</sup> For previous surveys of the behavioral approach to criminal law, see McAdams and Ulen Thomas (2009); Garoupa (2003); Jolls (2005). This survey differs, however, from these surveys as it aims also to explore the philosophical foundations of the field. For a general description of economic and behavioral approach to criminal law, see Harel (2012).



economics, the behavioral perspective is eclectic rather than unitary; it is composed of various psychological findings including cognitive biases and their effects, prospect theory, the effects of social norms, findings concerning the ways preferences and beliefs are being shaped, and even studies concerning happiness. Behavioral theorists call for the exploitation of various cognitive misperceptions, biases, and heuristics to increase the deterrent effect of criminal law prohibitions and sanctions and/or increase their effectiveness in other ways.

Both economic and behavioral approaches to law in general and criminal law in particular are revisionary. As the work of Gary Becker most famously indicates, even basic moral convictions deeply entrenched in legal doctrine are subject to scrutiny (Becker 1968). Yet, despite the principled willingness to challenge legal doctrine for the sake of promoting social goals such as efficiency or distributive justice, it is often the case that both economists and behavioral scientists use economic and behavioral tools to justify and defend, rather than critique and question, existing legal doctrines (Posner 2007).

THIS survey starts by examining in section 2 the theoretical foundations of behavioral analysis of criminal law. I contrast behavioral analysis with traditional criminal law doctrine (founded on retributive justice values) and, then, I contrast the behavioral approach to criminal law with traditional neoclassical economic analysis of criminal law and point out the distinctive features of the former. Section 3 illustrates the ways in which various behavioral phenomena can be used to understand the effects of criminal law norms and to design criminal law in a way that serves its social goals, in particular deterrence. Section 4 discusses the literature on happiness and its relevance to the optimal design of criminal law. Most interestingly I establish that unlike the conventional economic/behavioral analysis of criminal law, the literature on happiness can be used to promote retributive justice concerns. Section 5 examines critically the potential contribution of behavioral studies to the optimal design of the legal system.

## 2 THEORETICAL FOUNDATIONS

---

To understand the contribution of behavioral analysis of criminal law to the study of law one needs to point out what is distinctive about behavioral analysis of criminal law, namely in what ways behavioral analysis modifies the ways criminal law should be understood and/or reformed. In section 2.1 I contrast economic/behavioral analysis of criminal law with the traditional doctrinal/analytic approach based on retributive justice. In section 2.2 I contrast the behavioral approach with its older and more established ancestor—the traditional economic approach to criminal law—and examine the commonalities and the differences between these two fields.

## 2.1 Criminal Law versus the Economic/Behavioral Analysis of Criminal Law

This section explores the relation between behavioral analysis of law on the one hand and the dominant tradition in the study of criminal law, based on retributive values. It is not an accident that I speak here together on the economic and the behavioral analysis, as there are important commonalities between these two traditions. I have to concede at the outset that the retributive tradition is much richer than the version of retributivism presented below. Yet this sketchy version is sufficient to show the incongruities between traditional criminal law and the economic/behavioral tradition.

The traditional criminal law theorist believes that the criminal law primarily guides people and instructs them. Criminal law sanctions ought to be imposed on agents who committed wrongful acts because they “deserve” to be punished, and the severity of the criminal sanction ought to be proportionate to the wrongfulness of the act and to the culpability of the actor. (Duff 1990; 103; Fletcher 1978 454–59; Nozick 1981 363–97) Some retributivists oppose using criminal law for the sake of realizing any social goals including deterrence and/or just distribution, as such a use violates the basic Kantian principle under which one ought not use a person only as a means (not even as a means to deter or prevent crimes) (Rauscher 2012). It is unjust to inflict a sanction on the person simply because such a sanction brings about socially desirable outcomes; the only justification for such a sanction is that the person “deserves” it.

This view contrasts sharply with the view of law and economics and behavioral theorists. Under their view, criminal law is a mechanism or an instrument for preventing/detering undesirable behavior.<sup>2</sup> Most typically criminal law norms (as well as other legal norms) are perceived as incentives for individuals to behave in a way that is socially optimal. A state of affairs that is socially optimal is often identified with efficiency, but it need not be identified only with efficiency. Distributive justice concerns could also be regarded as a legitimate goal of the economic/behavioral analysis (Harel 1994, 1201–208; Harel and Parchomovsky 1999). Under this view, the primary role of criminal sanctions is to influence behavior (typically by deterring and sometimes also by incapacitating or rehabilitating criminals). Unlike the retributivist tradition, which often regards punishment as desirable in itself irrespective of its consequences,<sup>3</sup>

<sup>2</sup> This view follows the utilitarian theory developed by Bentham (1996). Gary Becker (1968) used contemporary neoclassical economic tools to develop Bentham’s insights. For a more legally informed doctrinal analysis of criminal law along these lines, see Posner (1985).

<sup>3</sup> The intrinsic value attributed to punishment by retributive thinkers is expressed most famously in a well-known passage in which Kant defends the view that if society is dispersed it must first execute all murderers “so that each has done to him what his deeds deserve and blood guilt does not cling to the people for not having insisted upon this punishment.” As the execution is performed just before society is dispersed, no social goals are served by it; it is executed exclusively because of the retributive concern to punish. See Kant (1996), 106.

economic and behavioral approaches regard punishment as evil in itself (given its costs to society and to the criminal) but, it is at times a necessary evil to deter or prevent crime.<sup>4</sup>

By regarding efficiency (or any other social goals) as the primary (or even exclusive) consideration underlying criminal law, economic analysis of law as well as behavioral analysis conflict with the retributivist tradition, and such a conflict inevitably triggers incongruities between criminal law as it is commonly justified and understood and the economic/behavioral approach to criminal law. Let me briefly explore two examples of such incongruity.

The first incongruity touches upon fundamental assumptions concerning human rationality. In different ways the traditional criminal law approach and the law and economics approach are founded on assumptions concerning rationality. In contrast the behavioral approach relies heavily on the irrationality of agents or, at least on assumptions concerning “bounded rationality,” namely, on the existence of limitations on rationality. Most typically, the behavioral approach to criminal law often calls for exploiting cognitive errors and irrational human dispositions to deter or prevent crime. Some of the recommendations of the behavioral tradition are designed to trigger errors on the part of the criminals in judging the severity of the criminal sanctions and/or the probability of detection in order to increase the effectiveness of criminal sanctions. The policy recommendations of behavioral theorists in such cases are founded often on methods that can be described as manipulative and fraudulent. For instance it was argued that to deter parking violations one ought to use “tricks” such as using “large, bright orange tickets that read ‘VIOLATION’ in oversize letters on the drivers’ side window where they are clearly visible to other drivers passing by.” The availability heuristic discussed below predicts that such a tactic would lead agents to overestimate the prospects of a parking ticket (Jolls, Sunstein, and Thaler 1998, 1538).

In contrast, classical retributivism is often based on claims concerning the rationality of individuals and their capacity to make informed moral judgments and act on their basis. Arguments concerning “free will” of individuals and their capacity to make autonomous choices are deeply entrenched in the retributivist tradition. This view has important implications concerning the justification as well as the legitimate ways of inflicting criminal sanctions. To illustrate, Anthony Duff’s “communicative theory” justifies punishment by pointing out that punishment conveys moral condemnation of the wrongful act (Duff 2007). The criminal process is described in his theory as a dialogue between the state and the criminal in which the state provides arguments and the criminal responds to these arguments. The use of manipulation by exploiting

<sup>4</sup> This was already the view of Jeremy Bentham (1996) who maintained: “all punishment is mischief, all punishment in itself is evil.” Bentham also inferred from this observation the principle of “frugality of punishment,” namely that punishment ought to be as small as possible to achieve its social goals. I believe it is unfortunate that this principle has been forgotten by contemporary legislators and judges.

cognitive errors (characteristic of the behavioral approach to law) deviates radically from the retributivist tradition, which has dominated the field of criminal law.

The second important incongruity between criminal law as understood by traditional criminal law theorists and the economic/behavioral approach to criminal law focuses on what counts as punishment. The retributivist believes that punishment ought to be inflicted because criminals “deserve” their punishment, and hence that what counts is the *ex post* sanction—the *actual punishment* inflicted on the criminal. The retributivist acknowledges of course that sometimes the criminal is not detected and, hence, no punishment is inflicted. But once detected the criminal ought to suffer in proportion to the gravity of the crime. In contrast, economic and behavioral theorists of law regard criminal law as an instrument designed to provide optimal incentives. The sanctions which are relevant for their enterprise are the *ex ante* sanctions—the *expected punishment* taking into account the probability of detection. Harsher actual sanctions are necessary therefore to the extent that the probability of detection is low and vice versa.

It follows from this analysis that distinctions that are central to legal doctrine such as the distinction between substantive criminal law, procedural law, evidence law, and the design of enforcement institutions are perceived by economists and behavioral scientists to be artificial. As the effectiveness of deterrence (as well as other social goals) hinges not only on the substantive doctrines of criminal law but also on the probability of detection and conviction, the law of evidence and the enforcement policy become central to the economic/behavioral analysis and are inseparable from the substantive doctrines of criminal law. One of the most interesting and somewhat counterintuitive results of this approach is that under both the economic and the behavioral approach, the optimal size of the criminal sanction is inversely related to the probability of detection and conviction. This view differs sharply from the retributivist tradition, which believes that the actual (rather than expected) punishment ought to be proportionate to the wrongfulness of the act and the culpability of the actor.

## 2.2 Economic versus Behavioral Approaches to Criminal Law

The standard law and economics account of criminal behavior begins with the observation that criminals (as well as other relevant agents such as judges, jurors, police officers, and victims of crimes) are rational decision-makers (Becker 1968; Posner 2007). Rationality however is understood differently than the rationality as understood within the retributivist tradition. Rational individuals as understood within this tradition are self-interested; they decide whether to commit a crime on the basis of weighing the expected costs and benefits resulting from it. These costs and benefits are not merely monetary; they include nonmonetary concerns including guilt feelings, stigma, and so on.

The traditional economic analysis of criminal law explains human behavior in terms of the expected costs and benefits of crime. These costs and benefits include parameters

such as the probability of detection, the size of the sanction, the attitudes of individuals towards risk, the expected costs of the sanctions, and so on. The basic premise of this analysis is that individuals make rational judgments on the basis of these parameters and guide their behavior accordingly.

Many of the behavioral theories examined in this chapter challenge this claim. For instance, it is pointed out that the expected sanctions do not guide people's behavior in mechanical or predictable ways. Criminal law influences individuals by modifying their beliefs and preferences. If there are systematic biases that distort the judgments of individuals, such biases alter individual behavior and result in irrational behavior. False beliefs concerning the severity of the sanction, the probability of detection, and so forth, would inevitably lead the criminal either to commit crimes it is irrational for him to commit or not to commit crimes it is rational for him to commit. Yet behavioral theorists also believe that such biases are not erratic or arbitrary; they are predictable and identifiable and therefore can be exploited by policymakers. At the same time, behavioral law and economics maintains that policymakers/legislators themselves are also subject to such cognitive biases, and those distort their judgments. Note that in the present context, the terminology of "biases" and "distortion" is not meant to be normative but purely descriptive; it is meant to denote that the behavior deviates (or is likely to deviate) from the assumptions of economic rationality.

The dichotomy between rationality and irrationality is not always precise or easy to draw. The controversy concerning rationality is complicated given that many of the distortions identified by behavioral scientists may be rational in the long run as they serve (at least in the long run) to promote the interests of the agents. They often reflect therefore a difference between rationality with respect to any individual decision and rationality in forming long-term rational strategies (Kahneman 2003).

The behavioral approach exploits empirical and experimental findings either to complement or, at times, to challenge the premise of rationality of human agents. The difference between complementing and challenging deserves attention. Behavioral scientists complement the findings by attributing to individuals certain preferences on the basis of psychological observations. Thus, for instance, behavioral scientists may establish that under certain circumstances individuals can be risk-averse or risk-loving depending on the way they frame a given choice or that they discount future benefits. These dispositions are not required by rationality, but they are not precluded by it either. At times behavioral scientists go further and challenge the findings of economic theory by pointing out that individuals are "irrational"; they form false or misguided beliefs that are not supported by the evidence at their disposal (e.g., they are too optimistic); they assess probabilities on the basis of anecdotal evidence; they make decisions based on the ways circumstances are being presented to them (framing) and not on the basis of how things really are. In such cases individuals operate in ways that fail to maximize their own utility.

The boundaries between complementing and challenging the findings of economic theory are not always clear, as it is not always clear what choices are rational or irrational. While it is always intellectually appealing for behavioral theorists to describe their

findings as refuting the rationality of the agents, the question of whether such findings complement economic theory (by identifying the actual beliefs and preferences of individuals) or establish that individuals are irrational (because their choices do not maximize their utility) is less crucial than simply identifying the behavioral phenomena and their potential relevance to legislators, administrators, and judges.

### 3 THE BEHAVIORAL APPROACH TO CRIMINAL LAW

---

Being ultimately a critical methodology designed to complement and challenge the findings of economic analysis of law, one traditional way of presenting behavioral findings is by contrasting them with the findings of traditional economic theory (Harel 2012). In the following discussion I shall follow this approach. At times however I will also contrast the behavioral findings with the retributivist tradition.

A word of caution: the behavioral analysis of law often applies general findings of behavioral science to the legal context. Psychologists investigate and make predictions as to how individuals act under uncertainty; what beliefs they form in different circumstances; what preferences they are disposed to adopt, and so forth. Legal theorists often apply these general predictions to the legal context. Such a methodology has risks, as predictions concerning the behavior of human beings are often sensitive to the context, and individuals facing a choice to commit a crime or to sentence criminals may behave differently than individuals facing choices in other contexts. To overcome this problem, behavioral law and economics theorists often examine empirically/experimentally the soundness of the general predictions in specific legal contexts. They do not apply automatically the general observations made by social scientists to the legal context. Instead they try to make independent investigations that involve *legal* uncertainty.

A much greater concern for the behavioral approach to law is the reliability of the behavioral method as such. Empirical/experimental research is currently a battlefield where different methodologies are being advocated and criticized, and theorists coming from different methodological schools expose deficiencies of other methodologies (see generally chapter 5 by Engel in this volume). This chapter does not examine these debates and it will use indiscriminately research by theorists coming from different schools, including in particular behavioral economics, cognitive psychology, and sociology. My primary aim here is to illustrate the ways behavioral approach can contribute to better understanding and designing of the law. The reliability of one method or another is of course important, but, as my aim is illustrative, it is not necessary to explore this issue here. The rest of this section examines various behavioral phenomena that are relevant to criminal law doctrine or related fields.

### 3.1 Behavioral Findings and Their Relevance to Criminal Law

As stated at the outset the behavioral analysis of law is an eclectic field. In this section I investigate various behavioral phenomena that are relevant to criminal law. The analysis is divided into two subsections. I first discuss behavioral phenomena that are individual and psychological (section 3.1.1), and then behavioral phenomena that are primarily sociological as they involve social interaction among individuals (section 3.1.2).

#### 3.1.1 *Psychological Findings and the Law*

##### 3.1.1.1. *Behavioral Observations on the Optimal Design of Criminal Sanctions and Probability of Detection*

Criminal law differentiates sharply between the size of the sanction and the probability of detection and conviction. The size of the criminal sanction ought to reflect the seriousness and hideousness of the crime. The more hideous the crime, the harsher the sanction ought to be. Murder is ordinarily a more serious wrong than burglary, and burglary is more serious than theft. The punishments for the different offenses should reflect the hierarchy or gravity of the offense. Under this view, there is no relation between the probability of detection and conviction, and the size of the sanction. Empirical studies indicate that the traditional strict separation between these two questions reflects not only legal doctrine but also the moral intuitions of most people. Individuals believe that the size of the sanction ought not to depend on the probability of detection (Sunstein, Schkade, and Kahneman 2000; Baron and Ritov 2009).

As indicated earlier, economic analysis of criminal law rejects this view. The primary purpose of criminal sanctions is to deter individuals from antisocial behavior. Increasing the probability of detection and conviction, and increasing the size of the sanction, are both congenial to deterrence. Both the probability of detection and conviction and the size of the sanction determine the size of the *expected sanction*, and it is the expected sanction that matters from the perspective of deterrence. As the expected sanction should not exceed what is necessary for the purpose of deterrence, it follows that the harsher the sanction, the lesser the probability of detection and conviction ought to be, and vice versa.

The legal system ought to determine not only the size of the expected sanction but also the size of the expected sanction's components: the probability of detection and conviction on the one hand and the size of the sanction on the other. In his seminal article on the economics of criminal law, Gary Becker provides a simple, compelling, and highly counterintuitive answer to this question (Becker 1968, 183–84). Under Becker's view, the answer to this question depends on the costs of increasing the size of the sanction on the one hand and increasing the probability of detection on the other. Becker maintains that if increasing the probability of detection is much more costly to



society than increasing the size of the sanction, it follows that the legal system ought to inflict harsh sanctions even for the most trivial offenses (Becker 1968, 183–84).

The possibility that efficiency may under certain plausible conditions require increasing sanctions and reducing the probability of detection horrified even the most orthodox advocates of law and economics who tried hard to provide counterarguments. For instance some theorists argued that harsh sanctions may have negative implications as they induce offenders to increase their investment in precautions, and therefore it may have negative effects on the probability of detection and conviction (Mikos 2005). Further it was argued that imposing harsh sanctions for all crimes undermines marginal deterrence; if I already committed a parking offense for which I am liable to be executed, I would not be deterred from committing more serious crimes, for example, killing eyewitnesses (Posner 2007, 221).<sup>5</sup>

The behavioral tradition addresses this concern differently. Behavioral scientists explore which components of the criminal sanctions have a greater deterrent effect: the probability of detection or the size of the sanction. More specifically, they argue that while both increasing the sanction and elevating the probability of detection affect criminal behavior, they need not necessarily have identical effects.

To clarify this point, let us define the concepts of risk-neutral, risk-averse, and risk-loving individuals. An individual is risk neutral to sanctions if he is indifferent as to two sanctions with equal expected value. The expected value of a sanction is the size of the sanction multiplied by the probability that it will be imposed. A fine of \$10,000 with a 1% probability of detection deters such an individual to the same degree as a fine of \$100 with a 100% probability of detection. An individual is risk averse if he is deterred more by a harsh sanction with a low probability of detection (e.g., \$10,000 with a 1% probability of detection) than by a light sanction with a high probability of detection (e.g., \$100 with a 100% probability of detection). A risk-loving individual is deterred more by a light sanction with a high probability of detection (e.g., \$100 with a 100% probability of detection) than by a harsh sanction with a low probability of detection (e.g., \$10,000 with a 1% probability of detection). If individuals are risk averse, the policymaker can increase deterrence by imposing harsh sanctions with low probabilities of detection; if individuals are risk loving, deterrence may be increased by imposing light sanctions with high probabilities of detection.

Criminology research has been struggling with the question what has greater influence on criminal behavior: the certainty or severity of the criminal sanctions (Grogger 1991). While empirical researchers debate this issue there are behavioral phenomena that support the claim that certainty should have greater effects than severity, namely that individuals are risk loving; they are deterred more by low sanctions with high probabilities of detection than by harsh sanctions with low probabilities of detection.

<sup>5</sup> I think that despite the fact that this is considered the standard and the most compelling reply to Becker's challenge, this explanation fails. Even if the sanctions are harsh, marginal deterrence can be guaranteed by differentiating the probability of detection for light and grave offenses, i.e., by investing greater effort in detecting grave offenses. For another attempt to rebut Becker's challenge along the lines of traditional law and economics, see Polinsky and Shavell (1984).

Take first the case of incarceration. Increasing the size of a sanction from a one-year to a two-year prison term does not double the deterrent effects of the sanction, because of a psychological phenomenon called discounting of the future.<sup>6</sup> Compare Arthur, who expects to go to the dentist tomorrow and have a painful treatment, with Betty, who expects to go to the dentist next month. Arthur is anxious and wakes up at night anticipating the pain, while Betty has no anxiety at this point. Individuals tend to discount the significance of distant future events. It follows that a second year of imprisonment (which inevitably starts only after the end of the first year) has a lesser deterrent effect than the first year (which starts immediately after conviction). A prison term of one year with a 2% probability of detection has a greater deterrent effect than two years with a 1% probability.

Another relevant explanation is based on the observation that legal sanctions are only part of the overall sanctions imposed on criminals. The criminal also suffers from stigma, which in turn often has both monetary and nonmonetary effects on the criminal.<sup>7</sup> Criminal conviction exposes the criminal to both legal and social sanctions. Assume that conviction exposes the criminal to a legal sanction of \$100 and to a social sanction worth \$100 to him. The sanction is effectively \$200. Assume also that the probability of detection is 1% and the expected overall sanction (consisting of the legal and nonlegal sanction) is therefore \$2. If the state increases the legal sanction from \$100 to \$200, the overall sanction increases from \$200 to \$300 and the expected sanction increases as a result to \$3. Doubling the size of the (legal) fine in this case from \$100 to \$200 does not double the overall sanction. Yet doubling the probability from 1% to 2% would double the expected sanction from \$2 to \$4. Increasing the probability of detection has a greater effect on deterrence than increasing the sanction.<sup>8</sup>

### 3.1.1.2 Behavioral Effects of Uncertainty: The Punishment and the Detection Roulettes

Criminal law tradition is committed to reducing as much as possible any uncertainty or unpredictability as to the scope of criminal offenses and the size of the criminal sanction. Such certainty and predictability is required by principles of the rule of law, and, consequently, such a principle is often entrenched in bills of rights and constitutions (Harel and Segal 1999, 281–85). On the other hand, there is no attempt on the part of the social planner to guarantee certainty with respect to the probability of detection or conviction.

<sup>6</sup> For an accessible explanation of discounting, see Shane, Loewenstein, and O'Donoghue (2002). For an application to the case of incarceration, see Harel and Segal (1999). Some theorists raised the conjecture that long periods of imprisonment have very small deterrent effect because of "hyperbolic discounting" (Garoupa 2003; Bronsteen, Buccafusco, and Masur, 2009).

<sup>7</sup> On the dramatic monetary repercussions of criminal conviction, see Lott (1992).

<sup>8</sup> Traditional law and economic theorists could easily accept such an analysis, and, strictly speaking, this observation ought not to be classified as "behavioral." I include it here as some behavioral scientists often emphasize the significance of stigma.

To illustrate, consider the following example. Arnold and Betty commit an identical offense under similar circumstances. Arnold is sentenced to 10 years, while Betty is sentenced to 5 years. This gap seems unjust and may perhaps provide grounds for appeal. There is no reason why different sanctions are imposed on individuals who committed identical offenses under identical circumstances. In contrast, assume that when Arnold commits the offense, police invest little in enforcement and, consequently, the probability of detection is low. The police then increase the investment in detection, and when Betty commits the offense, she is caught as a result of this special effort by the police. It is difficult to claim in such a case that Betty was discriminated against even if she could not have known when she committed the offense that the police would increase its investment in detection, and even if she can prove that she would not have been caught unless the police changed its enforcement policy.

Another indication of the difference between uncertainty with respect to the size of the sanction and uncertainty with respect to the probability of detection can be found in the information given to individuals. Criminal law provides information with respect to the size of criminal sanctions; it does not provide such information with respect to the probability of detection. Criminal law doctrine guarantees that the sanctions meted out would not be more severe than the one in force when the offense was committed, but it does not guarantee that the probability of detection remains fixed. It is a basic principle of criminal law (and it is part of the rule of law more generally) that increasing a sanction for a criminal offense does not apply retroactively. A potential criminal can “rely” on the size of a sanction as is specified in the law at the time she commits the offense. On the other hand, no legal system allows a criminal to argue that the probability of detection increased “retroactively” after the offense was committed.

More generally, different legal ethos governs the size of legal sanctions and the probability of detection. The severity of the criminal sanction reflects the seriousness of the offense; hence, the legal system is committed to consistency in determining the severity of sanctions. Most importantly, it is committed to providing “fair warning” to criminals with respect to the size of the criminal sanctions. The detailed Sentencing Guidelines are perhaps the most evident manifestation of the commitment of the criminal law system to provide a fair and precise warning. On the other hand, the probability of detection is under the dominant tradition a function of pragmatic considerations, which change from time to time. The legal system rejects punishment roulettes and tries to guarantee certainty and predictability with respect to the size of the sanction. It does not, however, oppose detection roulettes, and the probability of detection is subject to uncertainty.

The differential treatment of punishment on the one hand and probability of detection on the other hand appears natural to traditional criminal lawyers, but from an economic perspective it is puzzling. After all economic analysis of law regards both punishment and detection as components of the expected sanction. Why should there be such a major difference between the treatment of the size of the sanction and the probability of detection?

A natural way to justify the differential treatment is to explain it on behavioral grounds as an effective means to increase deterrence. This justification is based on the expected reaction of a criminal to punishment roulette on the one hand and probability of detection roulette on the other hand (Harel and Segal 1999). To illustrate, compare the two following legal systems. Under the first system, every convicted thief is sentenced to two years in prison. Under the second system, there is a sentencing roulette that inflicts a sanction of three years in prison on 50% of thieves and one year in prison on the other 50% of thieves. The expected sanction is two years in prison. Which system is more effective?

The answer to this question depends on the deterrent effects of each one of these systems. If thieves were risk averse, they would prefer the first system to the second system, and, consequently, sentencing roulette would have greater deterrent effect. If, on the other hand, thieves are risk loving, they would prefer sentencing roulette, and therefore the deterrent effect of a certain sanction would be greater. Similar observations can be made with respect to the probability of detection roulette. If criminals prefer probability of detection roulette (e.g., 50% of the criminals are caught with a probability of 1% and 50% with a probability of 3%) over a known probability (all are caught with a probability of 2%), the deterrent effect of probability of detection roulette would be lower than the deterrent effect of a known probability and vice versa.

We can now evaluate the desirability of sentencing roulette on the one hand and probability of detection roulette on the other. As we saw earlier, the existing legal system rejects the sentencing roulette but endorses the probability of detection roulette. The current system is justified only if criminals are risk loving with respect to sentences but risk averse (or at least risk neutral) with respect to the probability of detection.

There are indeed good reasons to believe that criminals are risk loving with respect to sentences. As demonstrated above, criminals are likely to be risk loving with respect to terms of incarceration because of their disposition to discount future costs. Hence, predictable (fixed) terms of incarceration (e.g., 2 years in prison) are likely to deter individuals more than risky terms with the same expected length (50% of 1 year in prison and 50% of 3 years in prison).

The infliction of fines requires a different analysis since unlike years in jail the entire fine is paid at once. Consequently there are no discounting effects. There is, however, theoretical support for the view that criminals are risk loving also with respect to fines. One of the major findings of prospect theory is that individuals are (typically) risk averse with respect to gains but risk loving with respect to losses. To determine whether a person is risk averse or risk loving, one ought first to identify whether the agent perceives the decision as involving a loss or a gain. The subjective perception of a decision as involving a gain or a loss is often highly sensitive to the way the decision is described to the agent and to the context in which the decision is being made. Thus, if individuals face a choice between an 80% probability of gaining \$100 or receiving (for certain) \$80, they would prefer receiving \$80. In contrast, if individuals face a choice between an 80% probability of losing \$100 or losing (for certain) \$80, they would prefer the lottery to the loss.

A fine is naturally understood as a cost, and, consequently, under prospect theory one would expect criminals to be risk loving with respect to a sentencing lottery involving fines. These observations support the existing legal regulation of uncertainty under which sentences are certain and predictable.

Yet, given the sensitivity of risk propensities to the “framing” of the decision as involving either a loss or a gain, one may reach also other conclusions. In an experiment designed to examine the risk propensities of individuals to criminal fines, it was found that the transition from a certain/predictable schemes of fines to a risky scheme of probabilistic fines increased rather than decreased the effectiveness of deterrence (Baker, Harel, and Kugler 2004). One possible explanation is that individuals do not evaluate the sanction in isolation; instead they evaluate it in conjunction with the expected benefits of the crime. In deciding whether to commit a crime individuals discount the costs (fines) from the benefits and, if the sum is positive, they treat their decision as a decision involving gains. Prospect theory would in such a case predict that criminals would be risk averse, and if they are risk averse the optimal sanction ought to be probabilistic.

This example illustrates that behavioral science predictions are often not well defined. Punishment is clearly a cost to individuals who commit crimes. But it cannot be assumed without further investigation that criminals treat punishment as a cost or that their risk propensities with respect to criminal sanctions are aligned with the predictions of prospect theory concerning losses. What determines whether a person is risk averse or risk loving is the framing of the decision by the agent as a gain or a loss and not any external or “objective” judgment as to its nature. As one theorist argued: “While the predictions of prospect theory are clear once a reference point has been established. . . it is far less clear what constitutes a reference point (Teichman 2011, 1700–701).<sup>9</sup>

### 3.1.1.3 *How to Enrich the State by Using Prospect Theory*

Tax evasion is among the most common criminal offenses and many resources are invested in an effort to reduce its scope. Behavioral scientists believe that prospect theory may be used to reduce the scope of tax evasion (Yaniv 1999; see also Guthrie 2002–3, 1142–45).

As mentioned above, prospect theory predicts that individuals have differential attitudes towards risk. Risk attitudes are different in cases in which the decision involves probabilistic gains and cases in which the decision involves probabilistic losses. While individuals are risk averse with respect to gains, they are often risk loving with respect to losses.

Tax evasion can be described as a lottery. The individual faces a choice to pay his taxes or to pay a smaller amount but to face a risk that, if caught, he would be subject to

<sup>9</sup> This is part of a larger concern raised by Teichman, namely the concern that some cognitive phenomena are indeterminate. I investigate this concern at greater length below.

a large fine. The inclination to take risks hinges on the question of whether individuals perceive the lottery as a lottery designed to increase their gains or to reduce their losses. Further, the state can influence (at least to some extent) whether the lottery is perceived as minimizing losses or maximizing gains. Thus, the state can partially control the risk attitudes of individuals and manipulate them to promote its ends, namely to reduce tax evasion.

One instrument used by the state is advance tax payment. The state deducts money during the year and, at the end of the year, the taxpayer is required to provide an annual report of his income. If the income is larger than the evaluation on the basis of which the advance payments were made, the taxpayer pays the difference to the state. If the income is lower than the evaluation on the basis of which the advance payments were made, the tax authorities pay back the difference to the taxpayer. Should the state make a high evaluation of the income (and therefore most likely return money to the taxpayer at the end of the year) or should it make a low evaluation of the income (and charge the difference from the taxpayer at the end of the year)?

Prospect theory would recommend that the state make a high evaluation. High advance tax payments mean that tax evasion is a lottery over gains rather than losses. The taxpayer has already made the payment and he expects to get a return that, it is likely, will be perceived by him as a gain. Given the prediction of prospect theory that individuals are risk loving with respect to losses, one may expect that individuals would be more inclined to engage in (risky) tax evasion under a scheme in which the advance payments are small (and therefore the lottery involves losses) than in a scheme in which the advance payments are high (and therefore the lottery involves gains). Deterrence considerations suggest therefore that the state ought to prefer a system in which advance payments are high over a system in which advance payments are low, as high advance payments will result in greater compliance with the law.

#### 3.1.1.4 *Prediction and Postdiction*

Much of the discussion so far has focused on uncertainty. One of the interesting findings related to decision-making in uncertain situations is the differential treatment of future versus past uncertainty (Guttel and Harel 2008). Psychological research suggests that individuals are less willing to bet on past events than on future events.

Assume that you have to bet on the result of tossing a die. In one case the die has already been tossed while in a second case the experimenter is going to toss it. It seems as if there is no difference between the cases. The probability of guessing correctly in both cases is identical. However, experimental research indicates that individuals react differently in these cases (Rothbart and Snyder 1970). In another famous experiment, subjects were asked to choose between two possible bets: one involved guessing whether a particular stock had increased or decreased in value on the day *prior* to the experiment and the second involved guessing whether a particular stock would increase or decrease in value on the day *after* the experiment. The results indicated that 70% of individuals preferred the second bet (Heath and Tversky 1991). It is shown below



that the social planner can use the differential attitudes toward the past and future uncertainty in order to increase the deterrent effects of criminal law.

Precautions against crime are divided into two types. Some precautions operate before the crime is committed (e.g., cameras and LoJacks). Other precautions operate after the crime is committed (e.g., police patrols). The empirical findings concerning uncertainty indicate that precautions of the first type are more effective than precautions of the second type. In the case of the first type of precautions, the criminal bets on precautions, which operate at the time *the offense is committed*. He is asked therefore to bet on a die that has already been tossed, for example, on the question of whether a camera documents his behavior. In the case of the second type of precautions, the criminal is asked to guess the probability of a future event, for example, a police patrol. The differential treatment of prediction and postdiction suggests that criminals are more likely to bet in the second case than in the first. Consequently, the first type of precautions is more effective.

One way to illustrate this point is to reexamine the operation of tax enforcement authorities. Typically, tax authorities use samples of individuals who are selected randomly. The sample is selected at the end of year. Taxpayers who consider committing fraud bet on the future; they bet that their names will not come up in the sample. It is easy to see how the system can change such that taxpayers bet on the past rather than on the future. If the lottery takes place not at the end of the year but at the beginning of the year, the taxpayers bet not on the question of whether their names *will* come up on the sample but whether their names *already appear* in the sample. This latter bet has greater deterrent effects.

Last, this analysis may have effects on the choice between the use of rules and standards. Both rules and standards are legal norms that adjudicators use to evaluate actions. Standards are open-ended norms, allowing the adjudicator to make facts-specific determinations, such as whether a driver used “reasonable care” in a given situation. A rule, conversely, is a more specific and concrete norm and, consequently, leaves less discretion to the decision-maker than a standard. The distinction between rules and standards is, in practice, a matter of degree. A legal norm can be more or less rule-like or standard-like.<sup>10</sup>

Rules and standards may both generate uncertainty. Standards are legal norms whose interpretation is provided only *ex post* by the courts. Standards, therefore, produce *future* uncertainty resulting from the indeterminacy of the interpretation given to them *ex post* by the courts. Rules are concrete norms that leave no (or little) discretion to decision makers. Yet individuals whose behavior is governed by rules are not always familiar with the specific details of a rule; especially where the applicable rule

<sup>10</sup> Note that the difference between rules and standards is not based on the semantics of the norm. The norm can be framed in vague terms and yet understood by decision-makers to be determinate and vice versa. My characterization here is not based on the literal meaning of the norm but on the way the norm is understood by those who are subjected to it and those who are assigned the task of interpreting the norm.



is complex, familiarization with its content is often costly. Individuals who know that their behavior is governed by rules—but fail to familiarize themselves with the rules—thus face uncertainty concerning already existing regulation. In contrast to standards, this type of uncertainty requires individuals to *postdict* the content of the law.

Let me raise here some conjectures following from these observations. Those are conjectures that have not been verified; instead they are based on extension of the general findings concerning prediction and postdiction. The experimental findings imply that individuals will be less inclined to engage in uncertain rule-governed activities than in uncertain standards-governed activities. The more rule-governed an activity is, and the more complex the rules governing it are, the lesser the inclination to engage in the activity. Transforming a standards-governed activity into a rule-governed activity transforms uncertainty (at least from the perspective of the individuals who are unable or disinclined to learn the rules) from future uncertainty into past uncertainty.<sup>11</sup> This change is likely to affect conduct. To better understand these effects, consider the regulation of driving and alcohol consumption. Drunk driving is currently regulated by rules rather than standards. These rules determine the upper legal limit of alcohol in drivers' blood. Although there are some attempts at producing a device that drivers can use to calculate blood alcohol content, most individuals do not and cannot determine what their blood alcohol content is. To this extent, from a driver's perspective, an alternative standard-like norm of regulating drunk driving—such as a standard that dictates a person ought not to drive when her ability to drive safely is significantly impaired by alcohol—would be as costly to comply with as the rule based on the blood alcohol content. Although the level of uncertainty and the cost of compliance are essentially similar under both forms of regulation, the behavioral findings suggest that drivers' behavior under a rule or a standard is likely to be different. More specifically, if the purpose of the legal norms governing drunk driving is to induce drivers to err on the side of caution, legal rules of the type used now may be preferable to legal standards of the type suggested above. Drivers who face legal rules based on blood alcohol content would be more risk averse than drivers who would face a legal standard prohibiting driving when, for instance, alcohol impairs significantly one's ability to drive.

### 3.1.1.5 *The Availability Heuristic and Criminal Law*

The traditional economic approach explores the influence of the size of sanctions and the probability of detection on deterrence. Behavioral economists argue that deterrence is not a product of the actual size of a sanction or the probability of detection

<sup>11</sup> Changing from a rule-based scheme to a standards-based scheme or vice versa is not a rare event. Speed limits could be specified numerically (a rule) or in vague terms such as "reasonable speed" (a standard). In 1995, Montana eliminated its numerical daytime speed limit on its interstate highways and adopted instead a standard of "reasonable and prudent" daytime driving. For a discussion, see King and Sunstein (1999). The standard eventually was invalidated on the grounds that it was unconstitutionally vague. See *State v. Stanko*, 974 P.2d 1132 (Mont. 1998).

but a product of the *beliefs* concerning the size of the sanction and the probability of detection. Can we examine how these beliefs are formed and shaped? Can we affect the content of these beliefs?

Some (and perhaps most) readers of this chapter have considered once or twice in their life whether to speed or to park illegally.<sup>12</sup> In such cases those readers also thought of the potential risks of such behavior: the risk of being fined. But (with the possible exception of the fine for illegal parking) it is likely that the readers did not know the precise sanctions for such behavior and certainly did not know the probability of detection.

How did those who decided to speed (or not to speed) or to park illegally and risk a fine (or drive for the third time around the block and look for a legal parking space) form their decision? There is perhaps one parameter that influences greatly such a decision. If on the evening before the event, one of your friends complained about getting a speeding ticket or you read in the paper a report on a police campaign against speeding, you are more likely to comply with the law. Psychologists call this phenomenon *availability*. The term “availability” denotes the disposition of individuals to form their beliefs on the basis of anecdotal information, which they can easily recall from memory (Tversky and Kahneman 1973). A famous example corroborating the availability heuristic is based on the following experiment. Individuals who are asked to estimate how many seven-letter words in a 2,000-word section of a novel end in “ing” give much larger estimates than individuals asked how many words in such a section have “n” as the second-to-last letter, despite the fact that objectively there are more words that satisfy the latter than the former. It is simply the case that individuals can more easily recall examples of the former type of words than the latter type of words. More relevant for us is the finding that people tend to overestimate vivid/salient risks, such as car and plane accidents, school shootings, and nuclear accidents, and underestimate less visible or publicized risks, such as heart disease. The former are well publicized, and therefore people tend to overestimate the prospect that they may occur.

Our beliefs concerning the size of sanctions and the probability of detection are not formed by reading the penal law or reading the annual statistics collected by the police. Empirical findings show that individuals have little information both with respect to the size of the criminal sanctions and with respect to the probability of detection (Robinson and Darely 2004). Instead these beliefs are often formed by a story we read in the news or an anecdote told by a neighbor.

Some theorists proposed to use the availability bias to reduce the rate of illegal parking by using colorful and visible parking tickets. The argument is that neighbors and pedestrians will remember such tickets, thus creating great deterrent effects (Jolls, Sunstein, and Thaler 1998, 1538). More generally, this view would imply that to be

<sup>12</sup> Interestingly thinking about the prospects of getting a parking ticket was the trigger for the seminal article on the subject of criminal law and economics by Gary Becker (Becker 1997). It seems that the only offenses that can excite the minds of professors are speeding or parking offenses.

effective, enforcement activity ought to be salient and vivid such that it will be registered in the minds of potential criminals. The availability bias also suggests that the public punishment used in the Middle Ages (e.g., public flogging or public execution) was congenial to deterrence not because it provided accurate information concerning the probability of detection, but because it provided a memorable and salient reminder to individuals of the risks of conviction. Imposing overly harsh sanctions and publicizing this fact may arguably be conducive to deterrence for this reason.<sup>13</sup>

It has also been pointed out that the availability heuristic influences not only potential criminals but public opinion and this may lead to suboptimal legislation or suboptimal law enforcement policies. One theorist argued, for instance, that “in criminal law, street crime (theft and violent crime) is especially vivid and frightful for most people. In contrast, white-collar crimes, such as financial frauds in which many victims lose small amounts, seem much less threatening” (Brown 2004, 342).

The availability heuristic may affect not only decisions by people subject to the law, but also the decisions of policymakers. Thus, interest groups may exploit the availability heuristic by using anecdotal evidence designed to affect public opinion. For instance, potential victims of crime may overinvest in precautions against crime due to the intentional manipulation on the part of firms expected to gain from selling such precautions. To do so, such firms need not lie about the frequency of crime; they simply need to publicize anecdotal horrific stories concerning crime.

### 3.1.1.6 *Overoptimism and Criminal Law*

One of the persistent findings of behavioral scientists is that individuals tend to be overoptimistic (for an overview see chapter 13 by Williams in this volume). For instance it was noted that individuals tend to believe that they are very unlikely to divorce even in the face of the statistics indicating a very high rate of divorce. It has been claimed that overoptimism weakens deterrence by causing potential criminals both to overestimate the benefits resulting from crime and to underestimate the probability of detection and the size of the sanctions (Garoupa 2003, 9). Overoptimism may also affect the behavior of victims of crime and cause them to underinvest in precautions against crime.

All these factors justify the conclusion that policymakers ought to take these factors into account in determining the size of the expected sanction, and impose a harsher expected sanction than the sanction sufficient to deter individuals who calculate correctly the expected costs and benefits of crime. One may doubt, however, the degree to which the optimism bias should be used by policymakers. As potential criminals often have no knowledge concerning the probability of detection, it is not necessarily the case that one can take the actual probabilities of detection as a starting point and infer that criminals' subjective assessments of the probability of detection are lower. Last, it

<sup>13</sup> Legal theorists have used this argument to justify the imposition of capital punishment. It was argued that given the salience of capital punishment, it would be highly effective (Sunstein and Vermeule 2005, 714).

was also pointed out that overoptimism may cause criminals to underinvest in precautions, that is, to be less careful; and such underinvestment on their part may be conducive to law enforcement (Garoupa 2003, 9).

### 3.1.1.7 *Positive Criminal Duties: The Duty of Rescue*

Criminal law typically consists of negative duties: it prohibits individuals from committing murders, thefts, and rapes. Liberal criminal law theorists are reluctant to impose positive duties. Yet this reluctance is not universally accepted. The approach of common-law systems on the one hand and European or religious systems on the other hand is different; the former systems are much more reluctant to impose a duty of rescue. Positive duties exist in criminal law even in common-law systems, but they typically hinge on the existence of prior relationships (such as the parent-child relationship) or special circumstances (such as drivers who observe a traffic accident). A classic case illustrating the common law's reluctance to impose positive duties is its refusal to embrace so-called Good Samaritan duties—that is, duties to rescue (Weinrib 1980).

Traditionally, the reluctance to impose positive duties is justified on grounds of autonomy. The legal system ought to protect the “negative liberty” of individuals but it cannot dictate to them what to do. William Landes and Richard Posner think that legal responsibility for failing to rescue characterizes communist or fascist legal systems, because the imposition of responsibility is a form of “conscriptio for the social service” (Landes and Posner 1978 420). Under this view, individuals ought to be legally required not to cause harm to others but they ought not to have any legal obligations whatsoever to help others. Can the absence of positive duties in the common law be justified?

Arguably, it is very difficult to explain the absence of positive duties on economic grounds. The utilitarian tradition, which provides the normative foundations for economic analysis of law, imposes very demanding duties on individuals. Maximizing utility requires one individual to help another as long as the marginal utility resulting from one's efforts is greater than the costs. My duty is therefore to serve the beggars of Jerusalem instead of sitting in my air-conditioned office and writing this text (Hills 2010). Naturally, it does not follow that the legal system ought always to impose such duties, as sometimes there are grave costs to legal enforcement. Yet it is quite difficult to explain why, if I sit on the beach watching birds while my desperate friend is struggling to save his life in the water, the law ought not to impose a legal duty to interrupt my favorite hobby and throw a rope to save him.

It is evident that often individuals engage in rescue even without a legal duty to do so (Heyman 2006). But it seems that imposing legal sanctions would increase the willingness to rescue. One explanation for the absence of such a duty provided by Landes and Posner, is based on the (highly speculative) conjecture that imposing such a duty would deter individuals from visiting areas where accidents may occur and, consequently, reduce the probability of rescue (Landes and Posner 1978, 418).<sup>14</sup> Another possible

<sup>14</sup> For a survey of the literature on the economic approach to the duty to rescue, see Harnay and Marciano (2009).

explanation is grounded in behavioral conjectures, and especially on the influential conjecture of Richard Titmus in his famous book, *The Gift Relationship from Human Blood to Social Policy* (Titmus 1997). In this book, Titmus identifies a psychological phenomenon that he labels “crowding out.” Titmus explores the practice of blood donations, comparing the American practice (in which blood donors receive monetary compensation) with the British practice (in which blood donors get no such compensation). Titmus found that the willingness to donate blood in Britain is greater than the willingness to donate blood in the United States despite the absence of monetary compensation in Britain. His claim (which is highly controversial) is that monetary compensation reduces or annuls altruistic incentives, and therefore, the blood supply in a society in which blood donors receive monetary compensation may be lower than in a society in which blood donors receive no such compensation.

Titmus focuses his attention on monetary incentives “crowding out” altruistic motivations, but his hypothesis can apply also to legal sanctions “crowding out” the same altruistic motivations. Under this view, by imposing a criminal law duty, law may weaken rather than strengthen the disposition of individuals to invest in rescue. Individuals may invest in rescuing precisely because they perceive it as a moral duty. Imposing legal responsibility for failing to rescue may turn the act from an act of charity, indicating the virtues of the rescuer, into an act that is merely done out of compliance with the law. Hence, legal sanctions may “crowd out” the altruistic motivations and thus reduce the willingness to rescue.<sup>15</sup>

### 3.1.2 *Sociological Findings and Criminal Law*

#### 3.1.2.1 *Criminal Law and Social Norms*

Social Norms Theory is based on the conjecture that there is an interaction between law and social norms. This view rejects the equation of criminal law sanctions as costs that decrease the inclination of individuals to commit a crime. Instead, the legal sanction itself influences individual preferences and social attitudes, and much of the influence of criminal law hinges on the resulting social pressures and stigma.

One branch of the social norms movement maintains that criminal behavior is not determined primarily by the size of sanctions or the probability of detection. A person’s criminal behavior is influenced to a larger extent by the behavior of other members of the person’s social group, the rate of compliance in the society as a whole, perceptions of the justness of the legal system, and so on (Posner 2000; Kahan 1997). The view under which law is merely an external incentive whose size is determined by legal sanctions does not reflect reality. In fact, there is an ongoing interaction between legal norms and social norms. The legal norms and the size of the sanctions inflicted on violators

<sup>15</sup> This conjecture may also be supported by Gneezy and Rostechini (2000). In this case it was documented that once a fine was imposed on parents who are late in picking up their children from daycare, the amount of late arrivals increased.

influence one's inclinations to perform the act and one's perception as to whether such an act is morally appropriate. The effectiveness of the enforcement of criminal law norms determines to a large extent the social attitudes towards the legal norms and, in particular, the social norms governing behavior.

A famous example identified with the social norms movement can illustrate these conjectures. The "broken windows" metaphor is used to convey the idea that the willingness of individuals to obey the law depends on their environment. In particular, the theory posits that minor violations—graffiti, abandoned buildings, garbage, and so on—regularly encourage criminal activity (Kahan 1997, 369). This conjecture led the former mayor of New York City, Rudy Giuliani, to strictly punish such minor violations, as he believed that individuals adjust their behavior not to the expected sanction but to the norms of behavior of their neighbors and friends (Harcourt and Ludwig 2006, 274).

In a famous experiment, the psychologist Phillip Zimbardo left a car with a broken window unattended and documented the resulting vandalism. Zimbardo found that the car had a negative effect on the behavior of individuals (Kahan 1997, 356). The influence of social norms has different explanations, some of which can be accommodated within the frame of neoclassical economics. One explanation is the "signaling" theory. Under this theory, individuals gain information from their environment with respect to the level of enforcement. Thus, minor violations (such as graffiti or broken windows) signal to individuals that the social order has collapsed and the probability of detection is low; therefore, crime is beneficial. Another explanation is based on the stigma effects of minor violations. If stigma is affected by the crime rate, a high rate for a crime indicates that there is no stigma attached to the crime.

### 3.1.2.2 *The Behavioral Effects of Just Legal Norms*

Traditional law and economics identify antisocial behavior as behavior that has negative externalities. One way to prevent antisocial behavior is to increase expected sanctions (either by increasing the probability of detection or by increasing the size of sanctions). One insight raised by behavioral scientists is that this ignores an important factor: the attitudes of individuals toward the law and, in particular, their beliefs concerning the justness of the legal system. Under this view there is an intimate relation between the belief that the law is just and the willingness to obey it. Perceptions concerning the justness of the law have motivational effects. The more a person believes that the law is just, the greater his willingness to obey the law. Justice Brandeis voiced this conviction when he said, "If the Government becomes a lawbreaker, it breeds contempt for law, it invites every man to become a law unto himself" (*Olmstead v. United States* (1928)). This is perhaps why some people also believe that justice should not only be done but also be seen to be done.

Nadler (2005) has demonstrated this point in a couple of experiments. In one experiment, she found that there is a correlation between an encounter that subjects had with a victim of injustice of the tax authorities and the willingness to evade taxes. In another experiment, some of the subjects were exposed to stories of injustice of the legal system

while others were exposed to stories that present the legal system as considerate and just. The first group showed much greater willingness to violate the law than the latter group.

These observations may have operative results. The decision to criminalize must take into account not only the negative externalities of the behavior but also the concern that criminalization undermines trust in the law. Criminalizing behavior that is perceived to be just may weaken the willingness of subjects to obey legal norms. Perhaps, for instance, the harsh treatment of immigrants that is perceived as unjust undermines respect towards the law in general (Kirk et al. 2012).

### 3.2 Summary and Critique

Section 3 provided various examples for the use of behavioral phenomena to understand the effects of legal rules and the effects of evidence rules and law enforcement policies. Further it also indicated how the lawmakers and policymakers can make use of these phenomena in designing laws and policies. Note that this section illustrates a central feature of the behavioral analysis of law, namely, that in contrast to traditional or classical law and economics, behavioral law and economics does not have a single unifying theory. It is based on numerous empirical and experimental findings. Applying those to the field of criminal law often requires sensitivity to circumstances and context, and should be done with caution. Mechanical application of psychological and sociological findings without examining their relevance to the criminal context is often misguided.

The primary accusation of behavioral scientists is that traditional advocates of law and economics blinded themselves to the realities of law and criminality. More specifically they argue that criminals are not self-interest maximizers, and do not operate in the ways attributed to them by economists. At best, we ought to complement traditional law and economics by examining what the real beliefs and preferences of criminals are. At worst, we ought to reject some of the premises of economic models, in particular premises concerning rationality. This chapter identified numerous behavioral phenomena that are relevant to the analysis of criminal law. These phenomena are only representative illustrations, and many more could be discussed.

I wish to devote this section to raising some critical comments on the behavioral movement and the applicability of its findings to criminal law. The behavioral analysis examined above is subject to criticisms of two types: internal and external. Among the internal criticisms, one may mention specifically what one theorist labeled “indeterminate biases,” namely the use of terms that acquire precise meanings only in specific contexts such as gains or losses in prospect theory (Teichman 2011, 1700–704). As illustrated above there is no natural way to classify punishment, and it could be classified either as a loss if looked at separately or as a gain if looked at in conjunction with the expected gains resulting from the crime. This is not unique to the case of punishment, and it raises doubts as to the potential contribution of prospect theory to



policymaking.<sup>16</sup> Further it was pointed out that the multiplicity of biases generates uncertainty as some of these biases may offset one another. People may for instance be overoptimists (and therefore underestimate the probability of detection) but, at the same time, be subjected to an availability bias that leads them to overestimate the probability of detection. It is difficult to predict under such circumstances which among conflicting biases is stronger or more effective (Teichman 2011, 1704–706).

Beyond these internal objections, there is a sense that behavioral law and economics treats individuals mechanistically. Punishment is designed to “train” the criminal. Concepts such as autonomy or choice, which are so central to criminal law, do not have a place within the behavioral tradition. Ironically, in the long run this view may erode the effectiveness of criminal law and, in particular, the effectiveness of the stigma attached to crime. If criminal law is nothing but a system of incentives whose effectiveness hinges on manipulation, fraud and cognitive biases (rather than a system of norms designed to guide individuals and aid them in deliberating on what ought and what ought not to be done), individuals would inevitably lose any feelings of shame or guilt or respect towards the criminal law. Instead, they would treat criminal law in the same way they treat powerful thugs. Such thugs inevitably intimidate, but their judgments do not guide individuals and their commands are disobeyed whenever it is safe to do so.

Last, some criminal law theorists believe that punishment is designed to cause pain to individuals and not only to deter them. Punishment is about retributive justice and it seems that retributivism has no place either in the classical law and economics tradition or in behavioral law and economics. The next part discusses the recent research on happiness, and examines its ambition to provide a better understanding of both deterrence and of retributive justice.

## 4 HAPPINESS AND CRIMINAL LAW

---

Jeremy Bentham argued already in the nineteenth century that happiness has noninstrumental (intrinsic) value, and that maximizing happiness ought to guide legislatures and policymakers.<sup>17</sup> Since then economists have abandoned the effort to measure happiness and turned instead to analyze well-being in terms of preferences (Crisp 2001, revised 2013). Yet Bentham’s view enjoyed a revival in recent years due to the work of psychologists, in particular the recent work by Daniel Kahneman (Kahneman and Thaler 2006). Kahneman distinguishes between two conceptions of utility: decision utility and experienced utility. Decision utility is based on the subjective evaluation of

<sup>16</sup> For an attempt to address this objection, see Zamir (2012, 889–92).

<sup>17</sup> For a discussion of the relevance of this body of scholarship to law, see Posner and Sunstein (2010).

the individual with respect to the future utility derived from his decision. Experienced utility, on the other hand, is the contemporary incarnation of Benthamite utilitarianism; it denotes the utility experienced or felt by the individual.

The research on happiness indicates that individuals do not always choose the course of action that maximizes their happiness. The experienced utility in “real time” differs from the predictions or beliefs of individuals concerning what is likely to bring them happiness; it also differs from individuals’ memories concerning the happiness they experienced in the past. In a now-classic study Philip Brickman and his colleagues compared recent lottery winners and recently paralyzed paraplegics and quadriplegics with control groups. Contrary to the predictions of many, the lottery winners were not much happier than the controls, and the accident victims were considerably happier than anticipated, reporting levels of well-being above the scale’s midpoint (Brickman, Coates, and Janoff-Bulman 1978, 920–21). A hungry client entering a food shop may purchase more food than he needs for the week simply because he is hungry (Kahneman and Thaler 2006). The actual utility derived from the food differs from the utility that the client predicts will be derived from the food. Individuals also overestimate the effects that changes in one’s life have on their experienced utility. When students from the Midwest and in California were asked questions about the happiness expected from living in these places, students tended to believe that life in California is much better than life in the Midwest. But when they were asked to indicate how happy they are in their own lives, it became evident that there is no gap between the happiness of students in California (as reported by them) and the happiness of students in the Midwest (as reported by them) (Schkade and Kahneman 1998). Thus, as Schkade and Kahneman (1998) note, “Nothing in life matters quite as much as you think it does while you are thinking about it.”

Happiness research may be especially relevant to criminal law theorists. Bentham well understood one reason for this. Criminal law in his view is a means of causing pain. But different people suffer differently from similar sanctions. Bentham, therefore, writes:

[O]wing to different manners and degrees in which persons under different circumstances are affected by the same exciting cause, a punishment which is the same in name will not always either really produce, or even so much as appear to others to produce, in two different persons the same degree of pain. (Bentham 1996, chap. 14)

Some contemporary legal theorists have argued that both retributivists and utilitarian theorists should take account of the differential effects of punishment on different people (Bronsteen, Buccafusco, and Masur 2009; Kolber 1990a, 1990b). Arguably, retributivists should care not about the punishment but about the subjective disutility experienced by the criminal; if this subjective disutility differs from one person to another, the sanction should also be different. The deterrence theorist believes that we ought to deter crime and that deterrence requires the imposition of a sufficiently large sanction. Such a sanction ought to be sensitive to the expectations of the criminal with respect to the disutility resulting from the sanction. It seems therefore that

happiness research is relevant to both retributivists and deterrence theorists. Yet the two approaches focus on different types of utility: the retributivists care about experienced utility (the mental states experienced by the person who is sentenced), while deterrence theorists focus on decision utility (based on the predictions of the individuals concerning their future utility) because the decision utility determines the deterrent effects of the sanction.

Let us explore first the retributive analysis. Psychological research concerning happiness can arguably provide new tools for retributivism. Under the subjective understanding of retributivism, retributivism should focus not on the objective size of the sanction (the size of the fine or the duration of the imprisonment) but on the subjective experiences of prisoners.

Assume that Mr. Sensitive and Ms. Nonsensitive are convicted of a crime and are sentenced to four years in prison. Mr. Sensitive suffers in prison while Ms. Non-Sensitive prefers freedom but is well adjusted to prison life. The fences, locks, and prison guards provide her with a sense of security and comfort, which she lacks when she is out of prison. According to the views of happiness theorists, retributivism ought to take into account the differences in their subjective experiences of imprisonment. Happiness research can contribute to the understanding and measuring of the subjective experiences of criminals and, consequently, also to calibrating sanctions in accordance with retributive justice considerations. For instance, Adam Kolber believes that the finding that rich people suffer more in prison than poor people requires (given retributivist assumptions) imposing harsher sanctions on poor people than on the rich. While this position is shared also by traditional law and economics theorists,<sup>18</sup> the reasoning underlying it is different. Kolber cares about the experienced utility of criminals rather than decision utility.

The retributivist hedonic view seems radical, but despite its radicalism, it has already some manifestations in contemporary legal practices. Some legal systems impose differential fines based on a criminal's income (Hamilton 2004). The rationale could be that identical fines impose greater disutility on the poor than on the rich. Judges often take into account in their decisions an offender's personal circumstances. This consideration by judges could be understood as being based in part on the view that the experienced disutility of an offender resulting from a sanction is sometimes too high and disproportional to the gravity of the offense.

One of the most surprising findings of happiness research is what is labeled "hedonic adaptation"—namely, the inclination of individuals to adjust to new circumstances such that dramatic changes in one's life do not radically change their happiness. For instance, individuals expect that receiving a large amount of money would greatly increase their happiness. Yet the research indicates that this is false. Similarly, individuals predict that their lives would be ruined by a serious disease or by being paralyzed. Yet the level of reported happiness after such a catastrophic event does not differ radically

<sup>18</sup> For references to the traditional economic literature, see Kolber (2009a, 232).

from the level of happiness before such an event. The typical reaction in such cases is a dramatic increase or decrease in subjective happiness, followed by changes that bring the level of happiness to that which was reported before the event took place (Kahneman and Thaler 2006; Brickman, Coates, and Janoff-Bulman 1978). Hedonic theorists of criminal law inferred that once a person enters a prison, one could expect a dramatic decrease in her happiness. But in the long run, the prison often has little effect on the happiness of convicted criminals. One of the implications of this observation is that the hedonic gap between short and long imprisonment is relatively small as people adjust to prison conditions (Merkel 2010, 229).

This finding is also important for an additional reason. When people are released from prison they seem to suffer long-term consequences, and these consequences affect their happiness. Further it was found that these long-term consequences are often not dependent upon the length of the incarceration (Bronsteen, Buccafusco, and Masur 2009, 1262–68). Even short periods of incarceration are likely to end in grave consequences. This poses a serious problem for any attempt to impose a proportionate sanction (understood in hedonic terms) as the difference between short and long term of incarceration is negligible or, at least, much smaller than expected.

While the hedonic analysis provided by criminal law theorists—and, in particular, its implications concerning retributivism—is intriguing, I believe that it is fundamentally flawed because it is based on a misunderstanding of retributivism. Retributivism is not based on the subjective experiences of the criminal but on the expressive significance of the sanctions as understood by the society as a whole. Punishment therefore ought to be understood objectively (on the basis of the duration of the imprisonment or the size of the fine or any other accepted objective measures) and not on the basis of the subjective disutility of the criminal (unless the subjective disutility could be conveyed to the public in the same way the size of the sanction can). Retributive practices are public, communicative, expressive practices designed to convey the intensity of moral disapproval as understood by society and not a practice designed to cause subjective disutility to the criminal.

Needless to say, if happiness research develops to the extent that would enable us to measure disutility and create a “happiness scale,” there is no reason why such a “happiness scale” cannot be used to determine the size of the sanction. In such a society judges could impose not ten years of imprisonment but 500 units of disutility. The directors of the prison could be in charge of enforcing this sentence, and different individuals would be subjected to different sanctions in accordance with the degree of subjective disutility resulting from the sanction. But to use such a system, what is crucial is not the subjective disutility as such but a public understanding of the retributive significance of units of disutility in the same way as the public understanding, in contemporary society, of the retributive significance of the size of fines and the length of imprisonment.<sup>19</sup>

<sup>19</sup> I have made this claim in the past (Harel 2012). Adam Kolber tries to address this objection (unsuccessfully in my view) in Kolber (2013, 1164–65).

Happiness research can be relevant also to deterrence theories, as it is relevant to evaluating the effectiveness of punishment in deterring criminals. Yet here, as observed above, what is crucial is not the experienced utility but the decision utility. In fact, deterrence theorists would be pleased if punishment could be predicted by criminals to be particularly harsh, but once it is imposed causes no disutility whatsoever. Deterrence is triggered by expected rather than experienced utility. Happiness research could contribute by shedding light on the displeasures of imprisonment (Bronsteen, Buccafusco, and Masur 2009, 1062).

Happiness research is young and is expected to develop further. To its credit, happiness research can struggle seriously with what is perceived by many to be the central role of criminal law—retributive justice. While utilitarians and economists provide no account of retributivism, happiness researchers use happiness as a tool to measure the severity of sanctions. But this research is based on a flawed understanding of retributivism: retributivism is not about subjective disutility but about the communicative and expressive significance of punishment. To achieve this purpose, fines and imprisonment are useful tools even if the disutility resulting from them differs from one criminal to another.

## 5 SUMMARY, CRITIQUES, AND FUTURE CHALLENGES

---

I will not deceive the reader by denying that the behavioral analysis of criminal law has so far had very limited effect on legal practice. There are very few fields in which economic analysis in general and behavioral analysis in particular have had lesser impact than in the field of criminal law.

Perhaps, as I argued in a different context, the reason is the great interest of the public in criminal law (in contrast to other more technical fields). Criminal law deals with murder, robbery, blood, and love, and beneath the gowns of judges one can sense intense passions and human sentiments (Harel 2012). The smell of blood, sweat, and sperm can barely be disguised when criminal law is at stake. Economics and behavioral economics seem too impoverished to govern this field where death, blood, sex, love, and hatred intermingle with each other. Perhaps philosophers rightly observe that retributivism is a primitive sentiment that ought to be overcome. But nobody has yet taught us how to do this, and the public and the legislature do not pay attention the pleadings of philosophers.

Furthermore, beyond the positive or negative effects of criminal law prohibitions on human behavior, it is still the case that the existence of criminal law prohibitions (independent of what they are or what their effects are) serves, as Durkheim observed, to reinforce social solidarity. Durkheim believed therefore that society needs crime. To illustrate why, Durkheim said: “Imagine a society of saints, perfect cloister of

exemplary individuals. Crimes, properly so called, will there be unknown; but faults which appear venial to the layman will create there the same scandal that the ordinary offense does in ordinary consciousness. If, then, this society has the power to judge and punish, it will define these acts as criminal and treat them as such" (Durkheim 1938, 68–69). Criminal law is not merely a means of training and inducing individuals to behave; it maintains and protects the social framework. This function cannot easily be translated into the language of economics or psychology.

These observations do not imply that economic or behavioral insights cannot be useful, but merely that their effects are at least ordinarily limited to the more technical aspects of criminal law, such as regulatory or white-collar offenses. Legal doctrine will continue to be governed by the Freudian Id rather than by the rational ideals of social scientists. I also dare say that this is not merely a prediction but also a hope. Criminal law is the field where the ideals of freedom and autonomy are particularly important. As mentioned above, the behavioral approach sharply conflicts with this view; criminal law is understood to be about training individuals to behave according to the norms rather than teach them about what is right and wrong and guide them in their moral deliberations. Such an approach undermines the pretense of the criminal law to guide us, to aid in deliberating and to provide an inspiration.

This is but an example of a gap in the literature on the behavioral analysis of law. More specifically I want to urge social scientists and legal theorists to think harder about what the normative significance of our preferences is. Precisely as in the context of criminal law I pointed out a tension between the ideals of autonomy and freedom of choice and the behavioral approach to criminal law, so such tensions can be found in other fields. Contract law theorists influenced by behavioral studies urge us to differentiate between our "true" preferences and those resulting from cognitive biases. Ultimately to know what we really want, we ought to launder the preferences and beliefs, to purify them. But the more successful behavioral scientists are in pointing out biases and misperceptions the less the faith one has in the very existence of independent and authentic preferences that merit respect.

## REFERENCES

- Baker, T., A. Harel, and T. Kugler. 2004. The Virtues of Uncertainty in Law: An Experimental Approach. *Iowa Law Review* 89:457–68.
- Baron, J., and I. Ritov. 2009. The Role of Probability of Detection in Judgments of Punishment. *Journal of Legal Analysis* 1:553–90.
- Becker, Gary S. 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76:169–217.
- Becker, Gary S. 1997. *The Economic Way of Looking at Life: The Nobel Lecture..* Stanford, CA: Hoover Institution on War, Revolution and Peace.
- Bentham, Jeremy. 1996. *An Introduction to the Principles of Morals and Legislation.* J. H. Burns and H. L. A. Hart, eds. Oxford: Clarendon Press; New York: Oxford University Press.
- Birke, R. 1999. Reconciling Loss Aversion and Guilty Pleas. *Utah Law Review* 1999:205–54.

- Brickman, P., D. Coates, and R. Janoff-Bulman. 1978. Lottery Winners and Accident Victims: Is Happiness Relative? *Personality & Social Psychology* 36:920–21.
- Bronsteen, J., C. Buccafusco, and J. Masur. 2009. Happiness and Punishment. *University of Chicago Law Review* 76:1037–64.
- Brown, D. 2004. Costs Benefit Analysis in Criminal Law. *California Law Review* 92:323–342.
- Crisp, Roger. 2013. Well-Being. In *Stanford Encyclopedia of Philosophy*. Available at <http://plato.stanford.edu/entries/well-being/#Hed>.
- Durkheim, Émile. 1938. *The Rules of Sociological Method*. S. A. Solovay and J. H. Mueller, trans., G. E. G. Catlin, ed. Glencoe, IL: Free Press.
- Duff, Antony R. 1990. *Intention Agency and Criminal Liability: Philosophy of Action and the Criminal Law*. Cambridge, MA: Blackwell.
- Duff, Antony R. 2007. *Answering for Crime: Responsibility and Liability in the Criminal Law*. Portland, OR: Hart.
- Fletcher, George. 1978. *Rethinking Criminal Law*. Boston: Little, Brown.
- Garoupa, N. 2003. Behavioral Economic Analysis of Crime: A Critical Review. *European Journal of Law and Economics* 15:5–15.
- Gneezy, U., and A. Rostechini. 2000. A Fine Is a Price. *Journal of Legal Studies* 29:1–17.
- Grogger, J. Certainty Vs. Severity of Punishment. 1991 *Economic Inquiry* 29:297–309.
- Guthrie, C. 2002–3. Prospect Theory, Risk Preference and the Law. *Northwestern University Law Review* 97:1115–64.
- Guttel, E., and A. Harel. 2008. Uncertainty Revisited: Legal Prediction and Legal Postdiction. *Michigan Law Review* 107:467–99.
- Hamilton, C. 2004. Making Fines Fairer. Available at <https://www.tai.org.au/documents/downloads/WP68.pdf>.
- Harel, Alon. 1994. Efficiency and Fairness in Criminal Law: The Case for a Criminal Law Principle of Comparative Fault. *California Law Review* 82:1181–229.
- Harel, Alon. 2012. Economic Analysis of Criminal Law: A Survey. In *Research Handbook on the Economics of Criminal Law*. Alon Harel and Keith Hylton, eds. Northampton, MA: Edward Elgar.
- Harel, Alon, and G. Parchomovsky. 1999. On Hate and Equality. *Yale Law Journal* 109:507–39.
- Harel, Alon, and U. Segal. 1999. Criminal Law and Behavioral Law and Economics: Observations on the Neglected Role of Uncertainty in Deterring Crime. *American Law and Economics Review* 1:276–312.
- Harcourt, B. E., and J. Ludwig. 2006. Broken Windows: New Evidence from New York City and Five-City Social Experiment. *University of Chicago Law Review* 73:271–320.
- Harnay, S., and A. Marciano. 2009. Should I Help My Neighbor? Self Interest, Altruism and Economic Analyses of Rescue Laws. *European Journal of Law and Economics* 28:103–31.
- Heath, C., and A. Tversky. 1991. Preference and Belief: Ambiguity and Competence in Choice under Uncertainty. *Journal of Risk and Uncertainty* 4:5–28.
- Heyman, D. 2006. Rescue without Law: An Empirical Perspective on the Duty to Rescue. *Texas Law Review* 84:653–738.
- Hills, A. 2010. Utilitarianism, Contractualism and Demandingness. *Philosophical Quarterly* 60:225–42.
- Jolls, C. 2005. On Law Enforcement with Boundedly Rational Actors. In *The Law and Economics of Irrational Behavior*, Francesco Parisi and Vernon L. Smith, eds. Stanford, CA: Stanford University Press.



- Jolls, C., C. R. Sunstein, and R. Thaler. 1998. A Behavioral Approach to Law and Economics. *Stanford Law Review* 50:1471–550.
- Kahan, D. 1997. Social Influence, Social Meaning and Deterrence. *Virginia Law Review* 83:349–96.
- Kahneman, D. 2003. Maps of Bounded Rationality: Psychology for Behavioral Economists. *American Economic Review* 93:1449–75.
- Kahneman, D., and R. Thaler. 2006. Utility Maximization and Experienced Utility. *Journal of Economic Perspectives* 20:221–34.
- Kant, Immanuel. 1996. *The Metaphysics of Morals*. Mary Gregor, trans. and ed. Cambridge: Cambridge University Press.
- King, R., and C. Sunstein. 1999. Doing without Speed Limits. *Boston University Law Review* 79:155–94.
- Kirk, D., A. Papachristos, J. Fagan, and T. Tyler. 2012. The Paradox of Enforcement of Law in Immigrant Communities: Does Tough Immigration Policy Undermine Public Safety? *Annals of the American Academy of Political and Social Science* 641, no. 1: 79–98.
- Kolber, A. 2009a. The Subjective Experience of Punishment. *Columbia Law Review* 109:182–236.
- Kolber, A. 2009b. The Comparative Nature of Punishment. *Boston University Law Review* 89:1565–608.
- Kolber, A. 2013. Against Proportional Punishment. *Vanderbilt Law Review* 66:1142–79.
- Landes, W., and R. Posner. 1978. Altruism in Law and Economics. *American Economic Review* 68:417–21.
- Lott, J. 1992. An Attempt at Measuring the Total Monetary Penalty from Drug Convictions: The Importance of Individual Reputation. *Journal of Legal Studies* 21:159–87.
- McAdams, Richard H., and Thomas Ulen. 2009. Behavioral Criminal Law and Economics. In *Criminal Law and Economics*, vol. 3 of *Encyclopedia of Law and Economics*, 2nd ed., Nuno Garoupa, ed. Northampton, MA: Edward Elgar.
- Merkel, D. 2010. Bentham on Stilts: The Bare Relevance of Subjectivity to Justice. *California Law Review* 98:907–88.
- Mikos, R. 2005. Enforcing State Law in Congress Shadow. *Cornell Law Review* 90:1411–85.
- Nadler, J. 2005. Flouting the Law. *Texas Law Review* 83:1399–441.
- Olmstead v. United States, 277 US 438, 485 (1928).
- Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Polinsky A. M., and S. Shavell. 1984. Optimal Use of Fines and Imprisonment. *Journal of Public Economics* 24:89–99.
- Posner, Eric. 2000. *Law and Social Norms*. Cambridge, MA: Harvard University Press.
- Posner, Richard. 1985. An Economic Theory of Criminal Law. *Columbia Law Review* 85:1193–231.
- Posner, Richard. 2007. *Economic Analysis of Law*. 7th ed. Austin, TX: Wolters Kluwer for Aspen Publishers.
- Rauscher, Fredrick. 2012. Kant's Social and Political Philosophy. In *Stanford Encyclopedia of Philosophy*. Available at <http://plato.stanford.edu/entries/kant-social-political/>.
- Robinson, P., and J. Darley. 2004. Does Criminal Law Deter? A Behavioral Science Investigation. *Oxford Journal of Legal Studies* 24:173–205.
- Rothbart, M., and M. Snyder. 1970. Confidence in the Prediction and Postdiction of an Uncertain Outcome *Canadian Journal of Behavioural Science* 2:38–43.

- Schkade, D., and D. Kahneman. 1998. Does Living in California Make People Happier? A Focus Illusion in Life Satisfaction. *Psychological Review* 9:340–6.
- Shane, F., G. Loewenstein, and T. O'Donoghue. 2002. Time Discounting and Time Preference: A Critical Review. *Journal of Economic Literature* 40:351–401.
- Sunstein, C. R., D. Schkade, and D. Kahneman. 2000. Do People Want Optimal Deterrence? *Journal of Legal Studies* 29:237–53.
- Sunstein, C. R., and A. Vermeule. 2005. Is Capital Punishment Morally Required? Acts, Omissions, and Life-Life Tradeoffs. *Stanford Law Review* 58:703–50.
- Teichman, D. 2011. The Optimism Bias of the Behavioral Analysis of Crime Control. *University of Illinois Law Review* 2011:1696–712.
- Titmus, Richard. 1997. *The Gift Relationship: From Human Blood to Social Policy*. Ann Oakley and John Ashton, eds. New York: New Press.
- Tversky, A., and D. Kahneman. 1973. Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology* 5:207–32.
- Weinrib, E. 1980. The Case for a Duty to Rescue. *Yale Law Journal* 90:247–93.
- Yaniv, G. 1999. Tax Compliance and Advance Tax Payments. *National Tax Journal* 52:753–64.
- Zamir, Eyal. 2012. Loss Aversion and the Law. *Vanderbilt Law Review* 65:829–94.

## CHAPTER 23

---

# BEHAVIORAL ECONOMICS AND THE LAW: TAX

---

EDWARD J. MCCAFFERY

### 1 INTRODUCTION: SITUATING THE TASK FOR TAX

---

THE extension of a behavioral approach of law and economics to tax law came somewhat late in the day (McCaffery 1994; Hill 2010; McCaffery and Slemrod 2006; Gamage and Shanske 2011; Galle 2013). On one hand, this is surprising, given the importance of tax to the economy of all developed nations, not to mention the deep political importance of the subject. Further, individual behavior has always been a central concern of public finance in the rational choice branch of economics, for policymakers need to pay attention to how individuals react to tax law changes in order to predict and respond to the revenue effects of reforms. Thus the field of optimal tax (Ramsey 1927) and optimal income tax (Mirrlees 1971), centerpieces of the traditional rational choice economic approach to tax, revolve around individual elasticities, or the changes to prices and behaviors brought about by tax law changes.

On the other hand, the late arrival of tax to the behavioral party is understandable, given the historic development of the field (Jolls, Sunstein, and Thaler 1998), and the challenges (which are still being worked out) of applying a behavioral approach to *public* finance (McCaffery and Slemrod 2006; Gamage and Shanske 2011). Behavioral law and economics first developed in the context of private law subjects such as contracts, torts, and property (Sunstein 2000; other chapters in this volume). In the public law context, early efforts looked at such specific matters as risk regulation (Noll and Krier 1990). This made sense because the wider field of behavioral economics arose as an alternative model of consumer or individual level *choice*, especially in situations involving risk (Kahneman and Tversky 1979, 2000; Thaler 1980). By framing and other

manipulations, ordinary subjects could be found to make inconsistent choices, which can lead to preference shifts or reversals.

The choices to be made in tax—to work or be unemployed, to save or spend, to give to charity or refrain from doing so, and so on—seemed in contrast to be straightforward, not about risk, and not necessarily the object of manipulation, because the government was more or less dictating tax law, without any individual-level negotiation in play. At first glance, framing does not appear to be relevant. Empirically minded public economists simply measured the facts of behavioral changes without pausing much to consider the reasons behind the changing behaviors; this is consistent with the traditional economics view of accepting individual preferences as is, and as being rational at least until proven otherwise. Rationality for these purposes means simply acting consistently with a well-defined utility function (Becker 1962; Sen 1977). Substituting leisure for work or consumption for savings in the face of increased taxes on work and savings seemed perfectly rational, and thus did not generate any grounds for deeper exploration.

In contrast, within more micro-level private markets where consumers spend most of their time, inconsistent individual-level decisions were easy to spot, once scholars started looking for them. Thus early applications of the *endowment effect*, *status quo bias*, or *loss aversion* featured tales and experiments of individuals valuing items, such as mugs, more if they owned them, or thought that they owned them, than if they did not (Kahneman, Knetsch, and Thaler 1990; Hoffman and Spitzer 1982). Damages in civil tort actions could easily be affected by *framing manipulations* (McCaffery, Kahneman, and Spitzer 1995; Kahneman, Schkade, and Sunstein 1998). Defaults in contractual terms could be sticky (Korobkin 2000). Government risk regulations were targeted at individual-level actions, and so it made sense to study them from a realistic behavioral perspective (Noll and Krier 1990).

Soon a paradox emerged and haunted the development of behavioral law and economics for some time. It seems abundantly clear, from both common sense and vast empirical observations, whether via controlled experiments or real-world observations (Camerer 1998), that individuals suffer from a variety of heuristics and biases in our individual decision-making: we all like our glasses half full, not half empty. We are not, that is, consistent. Still, the school of law and economics, acting as a subset of the rational actor economics tradition, was able to gain a good deal of traction ignoring all this and acting “as if” consumers were indeed rational (Posner 1972), even as counterindications continued to be found and to abound. Behavioral law and economics struggled to make specific policy prescriptions that could make any impact in the real world, and the whole endeavor was beset by charges of paternalism; specifically, of attempting to override consumer preferences, and violating the sacred consumer sovereignty principle (Sunstein and Thaler 2003; Camerer et al. 2003). Lingering tensions set the subject matter up for a seemingly never-ending war between two camps—one holding that individuals are indeed rational, the other that they are not. From a distance, one could see in this very struggle one of the most common biases: the all-or-nothing effect. Scholars and others assumed a binary choice between “rational” and “behavioral” approaches, and much ink was spilled and many trees felled hashing it all out.

Even at the time, however (and certainly now after decades of debate), the war of worldviews seemed a bit silly. Yes, rational actor models have been enormously helpful in understanding the law as it is and in prescribing reform, and we should be grateful for them. But yes, too, everyone one of us makes mistakes in reasoning all the time, measured simply on a within-subject basis as inconsistency.<sup>1</sup> How can this be?

The answer lies to a great extent on another side of the coin too often ignored in the early academic wars: the institutional and social structures in which decisions are made. Here there is a large difference between private markets and public finance—indeed, we might best refer to the public domain as a nonmarket one. Simply put, in private markets there are *markets* that serve as arbitrage and debiasing mechanisms. This is why Adam Smith (1776) saw markets, or the price system, as the “invisible hand,” and why Karl Popper (1945), in a wider social and political context, saw an “open society” as central to innovation and growth. As the US politician and president Abraham Lincoln put it, “You can fool some of the people all of the time, and all of the people some of the time, but you cannot fool all of the people all of the time.” Most of us need an escape, a curative, from our own biases and mental shortcomings at least a good deal of the time. Markets supply this, in large part because they give power to the few who are not fooled. Thus, the recent work questioning the endowment effect (Zeiler and Plott 2011) has mainly served to establish what should have been obvious all along: given a proper institutional design, individual-level inconsistencies in choices can be eliminated or at least significantly lessened.<sup>2</sup>

To illustrate, casually, I suspect that few if any readers know the marginal cost of supplying just about any product they purchase, such as a box of cereal (or the marginal cost of running a dishwasher, in a famous example from the literature [Liebman and Zeckhauser 2004]). It would not be difficult for a seller, in isolation, to trick us into thinking that the fair and accurate price was quite a bit higher than it is in reality. But in a fully competitive economy, sellers do not exist in isolation. In terms of pricing, competition helps to drive away or lessen any cognitive or behavioral advantage.<sup>3</sup> Marginal cost pricing, one of the central findings of the competitive general

<sup>1</sup> The field of behavioral economics has featured a wide array of methodologies, including: real-world observation, anecdotal reasoning, and experiments in the classic economics tradition featuring incentive-compatible stakes, using cash or other rewards to attempt to elicit true preferences, and comparing results across or between individuals. The initial (Kahneman and Tversky 1979) approach to behavioral economics was to use within-subject experiments or surveys (McCaffery and Baron 2006). In a within-subject design, the same person is examined against a null hypothesis that she should be consistent: if she likes a glass half full (or, in tax language, likes a child bonus) she should like it half empty (or a childless penalty), too. Behavioral economics has shown that this is not always the case (Thaler 2008).

<sup>2</sup> A series of subsequent studies have called Plott and Zeiler’s basic findings into question. (Knetsch and Wong (2009); Koh and Wong (2011); Isoni, Looms and Sugden (2011)). See Korobkin, this volume, for a discussion. I have no particular brief in this debate, although I have some confidence that some debiasing mechanisms can work. The point is rather that the kind of debiasing that at least sometimes obtains in private markets is altogether absent, or present in very different ways, in the public, nonmarket, realm.

<sup>3</sup> Once again, I am aware that this proposition is not without controversy on the private market side. According to Oren Bar-Gill, for example, in a competitive market, suppliers are expected to

equilibrium theory, obtains notwithstanding consumer's limited understanding—indeed total ignorance—of the marginal cost of inputs in just about any good or service we purchase every day. So it is still not completely clear that behavioral approaches to law and economics have all that much practical application where private markets exist to counteract pervasive bias.

Now let us turn our attention directly to tax, the subject matter of this chapter. And here, ironically, we can quickly see why a behavioral approach should be *more* powerful in regard to tax law than just about any other subject. After all, the central problem in tax and public finance is the almost complete absence of any institutional mechanism to offset or “arbitrage” against persistent individual-level cognitive bias (McCaffery and Baron 2006).

The point deserves emphasis. It is not biases in individuals' real-world reactions to a tax (for example, failing to appropriately reduce consumption of a good that is subject to a tax later imposed at the cash register, even where the information that the tax will be imposed is readily available [Chetty, Looney, and Kroft 2009]) that concerns me, as much as it is the failure of citizens to react consistently or rationally to tax law proposals put forth by politicians. Gamage and Shanske (2011) helpfully distinguish between “market” and “political” salience: the problem in this vocabulary is the low political salience of many real-world taxes. There are markets in politics, of course, but these markets are designed to get politicians elected. They do not necessarily give governments any incentive at all to debias the electorate. In fact, they tend towards the opposite. Thus, politicians favor hidden or low-salient taxes, such as corporate, payroll, or value-added ones, over more transparent ones, because they get government revenue at minimal psychic pain (salience) in order to fund highly salient expenditures. “Hidden” taxes have real effects, however, in terms of efficiency and on both “horizontal” (the norm of treating likes alike) and “vertical” (the norm that the rich or better-able-to-pay should pay more, in absolute and perhaps percentage terms) equity grounds. Thus we may—and, I have argued, have—end up with a tax system that is suboptimal, as measured by our own better-informed lights, in terms of both its social cost and its distributional properties (McCaffery and Baron 2005).

Taxes of any sort are hard to increase and attractive to cut, for obvious rational *and* behavioral reasons (McCaffery and Hines 2010). Government spending programs, the

compete over who exploits consumer biases more effectively. It seems especially likely that biases may affect non-price terms—not surprising, given that price is typically the most salient feature of a good or service in a private market, and behavioral economics strongly suggests that ordinary subjects have a hard time focusing on multiple dimensions. Bar-Gill and Board (2012). Yet once again, it is not my task to argue this matter out in this chapter—I mean rather more humbly to point to how the analysis, especially of the effect of arbitrage or debiasing mechanisms, should proceed very differently in private market domains versus public nonmarket ones. And, finally, I mean to stress that the project of analyzing the effect of nonmarket mechanisms on those biases in play in public finance systems is woefully lagging parallel research projects in private market domains.

other half of the fiscal equation, are conversely easy to raise and hard to cut, also for both rational and behavioral reasons. The net result, of course, is deficits of all sorts and at every governmental level. This is a perfectly predictable outcome of pervasive cognitive biases, playing out in a setting almost wholly lacking in institutional mechanisms to offset them.

It would seem, then, that the normative task for a behavioral law and economics approach to tax is to come up with ways to debias or counteract widespread behavioral biases and their effects, as evident in broad-level questions of tax system design. Such a project, however, has barely begun, and the early efforts in this direction, discussed below, seem tentative at best. In the meantime, advances have been made in other sub-fields, as scholars look to pluck the low-hanging fruit first.

The two major sets of “practical” recommendations to emerge thus far in tax and behavioral law and economics literature, however, suggest more that caution needs to be used when adapting behavioral economic findings to real-world public policy contexts than anything else. One application is to advocate the use of tax or otherwise favored savings plans to help cure persistent individual-level myopia. This suggestion suffers from a lack of a wider context and deeper analytic understanding of tax: when the facts of tax are better understood, one can see that the “cures” for myopia are likely to be undercut by . . . myopia.

The second application is the policy recommendation that governments should choose taxes with low “market salience,” and take steps to keep their “political salience” low. This line of reasoning suffers both from a certain obviousness—politicians, if they know anything, know *this*—and from a naïveté about the real-world effects of even hidden taxes, and about the possibilities of counteracting these effects by other tax law changes. Both of these examples are discussed at greater length below. They show us that there is still a long way to go from observation, and even experimentation, to prescription when it comes to the relevance of behavioral economics to tax.

The balance of this chapter first considers the more specific subjects of tax compliance and using tax as part of social behavioral modification. Afterward it turns to the relevance of behavioral approaches to tax-system design and, finally and most particularly, to the increasingly urgent question of fiscal balance.

## 2 COMPLIANCE MATTERS

---

Given the historical development of behavioral law and economics discussed above, it is not surprising that one of the earliest and still most developed applications to tax has been in the matter of tax law compliance. Whether or not to pay one’s taxes, especially in a largely self-reported tax system such as the income tax in the United States (and worldwide), is a choice, and one that involves a certain amount of risk—classic territory for a behavioral law and economics approach.



At the core of the compliance issue seems to be a rational choice calculus: comparing the marginal cost of compliance (paying all of one's taxes when owed) with the marginal benefits of noncompliance (taxes saved minus the penalties for noncompliance). In equation form, a rational individual, concerned only with dollars and cents in his pocket, should pay his tax,  $t$ , only if that tax is less than the sum of the probability of getting caught and forced to pay the tax,  $p$ , times the tax otherwise owed, plus the costs of any government fine or penalty for noncompliance if caught, plus the cost (if any) of noncompliance,  $c$  (I ignore for simplicity the costs of compliance):

$$t \leq pt + (1 - p)0 + pf + c, \text{ or} \\ t \leq p(t + f) + c$$

The problem is that, by just about any light, there is far “too much” compliance with the law, and far “too little” government commitment to either increasing  $p$ , by devoting more resources to tax law compliance, or  $f$ , by increasing the penalties for noncompliance (Alm, Vazquez, and Togler 2010; McCaffery and Slemrod 2006). Does this mean that there are behavioral biases in play?

Perhaps, but this would seem to be an area where new approaches also have their limits. The answer to the puzzle of excessive private compliance and suboptimal government enforcement, if that is what it is, seems to lie in the fact that there are other things going on. Politically, aggressive tax enforcement, at least of highly salient taxes, such as the personal income tax in the United States, is unpopular. Also, compliance does seem to turn on whether or not the government has mechanisms for checking on self-reporting (Slemrod 2007). There is considerably more compliance where third-party reporting exists to aid the government, as where employers must report wages paid, by law and in order to get their own tax law deduction for salaries paid, and employees are told what is reported to the government. In contrast, in cases where such third-party reporting does not obtain, such as sole proprietor income, noncompliance is far more common. This is all perfectly rational.

Nonetheless this field has been studied, and research in the area is growing of late. Scholars look to whether patriotism affects compliance, for example, or if other framing manipulations might be in play (Konrad and Qari 2012). Bruno Frey (2007) explains compliance within the framework of basic deterrence policy, in which people consider tax evasion in terms of the probability of detection and the degree of punishment. But Frey also suggests that forces such as the perceived behavior of others and whether the taxpayer believes that she received a “fair share” of public goods can affect compliance (Frey 2007).

Still, “overcompliance” may be due to risk aversion, or to a sense of duty, neither of which seem “behavioral” in the sense of being irrational. Granted, there is almost certainly widespread cognitive error when it comes to calculating the costs and risks of noncompliance—actually putting numbers to the compliance equation set out above. Yet although it is clearly a worthwhile social project to study and better understand the ways in which compliance with tax laws might improve, there seems something at least

a bit untoward in attempting to “debias” individuals from their “excessive” obedience to the law.<sup>4</sup>

### 3 BEHAVIORAL MODIFICATION

Another category of application of behavioral economics to tax law has been to embrace the fact that taxes can affect behavior, and then to look to see if tax can be the “answer” or corrective to certain behavioral biases perceived as being harmful to individuals.

Many biases cluster under the heading of *time-inconsistent* preferences: people do things today, failing to account for their longer term effects, and then later come to rue the day that they neglected the now-present future (Loewenstein and Elster 1992). Note that this is distinct from the more general topic of “sin taxes.” Taxes on goods or activities with negative externalities, and negative effects on those who consume them, such as cigarettes or alcohol (or under President Obama’s healthcare plan, on not being insured), after all, fall within the “rational actor” side of things: people will do less of any activity if its price is higher, and taxes can raise the price of an activity. This idea harks back at least as far as Pigou (1920), although modern public finance scholars have employed more sophisticated models, using behavioral insights to supplement the neo-classical perspective (Gruber and Köszegi 2004).

The behavioral approach adds that individuals, on their own lights, may be failing to optimize because of time-inconsistent preferences. Smoking and any other self-harming activities can fit the bill: the self who is doing the harmful thing is irrationally (inconsistently) discounting the future self who will suffer the consequences. In the tax law context, the canonical example has been the failure to save, which could equivalently be called overconsumption. “Rational” models of savings suggest a smoothing out of consumption over one’s life cycle (Modigliani 1966), and perhaps even an extremely or hyper-rational savings across generations and in reaction to government fiscal policies (Barro 1974). Behavioral economists have countered with a “behavioral life-cycle” model in which, lo and behold, most people most of the time fail to do the optimal thing, and make a wide range of mistakes in their saving and spending decisions (Shefrin and Thaler 2007). Simply put, most people seem myopic in their

<sup>4</sup> The topic of compliance to tax laws raises an array of other issues that I set aside in the interests of space. There are, for example: (a) Issues relating to when non-compliance is more likely, such as the difference between people who owe money on their taxes and those who are entitled to a refund, Robben et al. (1990); (b) Different ways to bolster deterrence using cognitive biases that are specific to taxes. See, for example, Guttel and Harel (2008), considering the “audit lottery” in tax enforcement; and (c) Recent findings relating to behavioral ethics and motivated reasoning that might also be relevant to tax compliance. For example, Feldman and Teichman (2008) argue that legal uncertainty in the tax code might generate a perverse effect with respect to compliance; see also Scotchmer and Slemrod (1988). The topic of the behavioral dimensions of tax law compliance could easily fill a separate chapter or even volume, and I apologize for the brevity here.

savings decisions and therefore they undersave—again, in a behavioral economics vein, as measured by their own *consistent* preferences over time. The idea has led to perhaps the poster child for a behavioral approach to public policy: tax-favored and highly salient savings plans, such as individual retirement accounts (IRAs) under US law, and/or provisions requiring employees to default into employer-provided retirement plans, relying on the status quo bias (or, more simply, inertia), to keep them there (Thaler and Benartzi 2004; Choi et al. 2002; chapter 12 by Korobkin in this volume).

But there is a problem with this particular set of policies that points to a more general danger with a behavioral approach to public policy (McCaffery 2006). As noted above, behavioral biases exist in context. Generally speaking, institutional forces such as competition or markets can ameliorate biases. Other times, in contrast, aspects of institutional design can undercut a behavioral patch. The case for making pension or other retirement plans both tax-favored and defaults turns on the fact that people suffer from *myopia*, an excessive concern with the present (and a corresponding devaluation of the future). The idea is to use one bias, the stickiness of the status quo, to offset another, *myopia*.<sup>5</sup>

To be fair, the strategy of making savings easy, salient, and sticky might work. The idea certainly makes sense in the abstract. The problem is that we do not live in the abstract. Aspects of institutional design—in this case, the analytic structure of tax—do not go away because we are not looking at them; indeed, the most general behavioral bias can be seen as an *isolation* effect, in which individuals, including policymakers, ignore facts and factors seemingly offstage or not salient at the time another decision is being made.

More specifically with tax-favored savings plans, the problem begins with the fact that the tax system as presently structured has a large bias in *favor* of present consumption, and *against* savings (McCaffery 2005). A major aspect of this structural bias is that debt, or consumer borrowing, is not taxable “income” when incurred. A US taxpayer using a “traditional” individual retirement account (IRA) is thus able to borrow, tax-free, in order to come up with enough cash to fund a contribution to a retirement plan. Today, this generates a tax deduction (often literally today, as the deadline for making a tax-deductible IRA is generally the due date of the return, i.e., April 15).

A taxpayer, such as a hypothetical Jane, can reduce the taxes she must pay within minutes by this trick. Jane borrows \$2,000 on a credit card. She then puts that \$2,000 into an IRA or other tax-favored vehicle, all on the same day, April 15, also known as Tax Day. Why would Jane do this? Perhaps because she is *myopic*: the very behavioral trait leading to the policy recommendation in the first place. The moves allow Jane to reduce the taxes she owes, *today*, by perhaps \$600 if she is in a 30 percent marginal tax bracket. The taxes Jane would otherwise owe fall by \$600, or her refund increases by that much. She has, today, another \$600 to spend, today. Now there is no free lunch

<sup>5</sup> The chapter by Pi, Parisi, and Luppi, found in this volume, discusses the issue of using one bias to offset another.

here—the \$600 tax bill lies in wait, until Jane withdraws the \$2,000 from her IRA. But that withdrawal, and its tax obligation, will not come today. So Jane doesn't think much about that ultimate cost. Why? Because she is myopic. Now, how, you might ask, could Jane possibly know of this seemingly sophisticated arbitrage technique? Perhaps because the financial institutions that collect fees from IRAs and interest from consumer loans have told her about the idea—again, we must always consider the institutional setting, here a private market, in which individual actions take place.

Note that in this example, wherein Jane borrows with one hand and saves with the other, there is *no* net savings; the positive savings in the IRA (\$2,000) are offset by the borrowing outside of it (−\$2,000). Indeed, if Jane goes forth and immediately spends the greater tax refund she gets (as the behavioral concept of mental accounts [Thaler 2008] and as myopia suggests that she will) there is *reduced* aggregate savings. Further, taxpayers like Jane fail to account for the taxes implicitly due from their retirement accounts, though these are inescapable, even after death.<sup>6</sup> A kind of *money illusion* leads people to think that they are richer than they actually are because of their ignoring of the built-in tax liability. This could easily lead to *less* saving going forward. Admittedly, this is speculative, based on a conjoint consideration of the analytics of current tax law design and familiar behavioral biases (McCaffery 2006). But some confirmation can be gained from the fact that the costs of the pro-savings provisions, in the form of foregone revenues (what tax experts call “tax-expenditure” analysis) seem to *exceed* the amount of new savings actually being produced in the United States each year (Bell, Carasso, and Steurle 2004).

This relates back to a previous warning. The point I am making is not a concern over paternalism, or about which preferences to count when individuals are inconsistent over time. It is simply that ad hoc patches to behavioral biases may not work, precisely because they are ad hoc. Water finds its own level. A persistent myopia will lead to “too little” savings, until and unless institutional structures are properly designed to counteract the bias. Ad hoc cures can be circumvented and only serve to make the web of tax more tangled, and the problem of undersaving worse.

## 4 TAX LAW SYSTEM DESIGN

---

So far this chapter has considered that behavioral approaches might be limited in their practical applications for tax law compliance and for curing “problems” such as too little savings through ad hoc patches. These happen to be the two most common domains

<sup>6</sup> A traditional IRA under United States law is funded with pre-tax dollars. IRC § 408. If a taxpayer holds an IRA until death, either the ultimate beneficiaries will pay ordinary income taxes when they receive distributions, or the decedent's estate itself will pay tax, as “income in respect of a decedent,” or IRD. IRC § 691.

where behavioral approaches to tax law have been invoked thus far. Does that mean the task for a behavioral approach to tax law is limited?

No. Getting back to my initial comments, all or at least many roads of a behavioral economics approach to tax seem to lead to a path less taken: considering the relevance of behavioral perspectives to the broad question of tax law design. Here the various mechanisms of arbitrage that lessen and perhaps eliminate the effects of bias—or, at a minimum, affect how biases play out—are almost altogether lacking. Politicians get elected based on pleasing the electorate, and hearing about taxes is rarely pleasing. How does widespread cognitive error play out in the nonmarket domain of public finance? This is a large and challenging project, which naturally breaks down into positive and normative dimensions.

## 4.1 Positive Findings

Let us begin with some positive or simply descriptive findings about popular perceptions of tax.

It is hardly surprising that people do not generally like taxes. There is thus a *tax aversion*, related to the more general behavioral bias of *loss aversion*. (See generally chapter 11 by Zamir in this volume.) In sum, people react to a perceived loss more negatively than the “mere” failure to obtain a gain, even of the same absolute magnitude, and even with the same end points. Thus, famously, Thaler (1980) observed that customers would use cash to avoid a penalty for using credit cards (\$3.90 a gallon versus \$4.00, say), but would use credit cards and forswear a “bonus” for using cash (\$4.00 versus \$3.90). Such biases would seem to extend to the purely formal matter of how a government exaction is labeled, whether as a “tax,” a “surcharge,” a “user fee” or something else. Although experimental evidence does not cut simply in any one direction (McCaffery and Baron 2006) it seems clear that the label “tax” alone creates negative attitudes.

Tax laws in advanced democracies are complex and have many dimensions. For broad-based comprehensive taxes, such as the income tax in the United States, which rely on a pattern of progressive marginal tax rates, there is a persistent confusion between marginal and average tax rates (see illustratively Liebman and Zeckhauser 2004). Thus individuals may react to the highest marginal tax rate and judge a tax accordingly. A somewhat related bias—at least a similar failure to understand the math of the matter—is the *metric effect* (McCaffery and Baron 2003). Simply put, citizens favor more progressivity when taxes are stated in the metric of percentage rather than in absolute dollar terms, as even a flat-rate percentage tax generates an appearance of progressivity when stated in absolute dollars. A flat rate 25 percent income tax means that a family making \$50,000 pays \$12,500 in taxes, and one making \$200,000 pays \$50,000. Comparing \$50,000 to \$12,500 makes the system look progressive, though, in a “better” metric, looking to average tax rates, it is not.

More generally, individual judgments about tax seem highly vulnerable to *framing effects*, or preference reversals based on the purely formal presentation of a choice set (Kahneman and Tversky 1979). A simple example is that individuals support “child bonuses” in the tax law, but recoil from the idea of “childless penalties”—although these are, analytically, the same thing, just different ways to put the same point, measured off of different baselines (McCaffery and Baron 2004). Thus, the tax system could tax couples at \$5,000, but give a \$1,000 tax break (“bonus”) to couples with children, or the system could tax all couples at \$4,000 and “penalize” taxpayers without children by \$1,000. Either way, a couple with children pays \$4,000, and a couple without pays \$5,000. Yet the “bonus” frame is widely liked whereas the “penalty” one is widely disliked. Progressivity, another dimension of many comprehensive individual tax systems worldwide, exacerbates the problem, as Schelling (1981) first speculated. Thus in a system without accommodation for children in the baseline, individuals support adding on child bonuses that are higher for the poor than for the rich. But in a system with accommodation for children already factored in, individuals support childless penalties surcharges that are higher for the rich. That is a double inconsistency: supporting progressive bonuses while opposing regressive penalties (McCaffery and Baron 2004).

*Salience* is key to many behavioral effects: individuals overreact to highly salient matters, and underreact (or fail to react at all) to low-salient ones. The concept of salience has led to a small explosion of scholarly work of late. The seminal piece in this field is the real-world experiment staged by Chetty, Looney, and Kroft (2009), in which the authors found that consumers had little if any reaction to taxes applied to store goods when the tax was added on at the cash register, as opposed to stated on the sticker price on the store aisles, even when the fact that the taxes would be later applied was made readily available to them (Chetty 2009; Galle 2013; Gamage and Shanske 2010, 2011; Schenk 2010). In a behavioral (or commonsensical) vein, this finding is hardly surprising, confirming individual predilections for “hidden” taxes, the more hidden the better (McCaffery 1994). It is the normative and public policy ramifications of this finding that still need development, as discussed in the next subsection.

There are still more behavioral biases at play in tax. A failure to add is central to many biases, such as *subadditivity*. In tax, individuals seem more willing to accept a set of smaller taxes than a single, large, tax, even where the former add up to more than the latter. (The seminal paper on subadditivity in general is Tversky and Koehler [1994].) This bias works hand in glove with issues of salience, for a series of low-salient taxes produces less psychic pain than an equivalent, large, transparent tax. People also generally fail to account for the deficiencies in one tax—measured by their own preferences—by adjusting another. Thus, people do not always act consistently when a single tax system is broken into pieces: if one tax on wages, such as the payroll tax, is considered to be too flat or even regressive, subjects still underadjust another tax, such as the income tax, to accommodate for the perceived shortcomings (McCaffery and Baron 2003). Taxes can also become bundled with spending programs, with a similar *disaggregation bias* in play. Thus if the government ceases to provide a certain good or service, leading, *ceteris paribus*, to a reduction in taxes, individuals have difficulty adjusting the tax cut



to maintain the same degree of progression that had obtained when the publicly provided good was bundled with the tax (Baron and McCaffery 2006).

Suppose, to illustrate this latter point, the government provides some benefit—schools, roads, healthcare—worth an average of \$5,000 per household. The benefit is financed with a progressive tax that, to keep things simple, imposes a \$2,500 burden on the poorest half of households and a \$7,500 burden on the richest half. If the government stops providing the good and eliminates the tax funding mechanism, not only will the good no longer be publicly provided, but also the cross-subsidy will end. If households need to come out of pocket to provide the good or service for themselves, the poorer half will “lose” \$2,500 because the cost to them will double; the richer half will “win” the same \$2,500 by saving on the cost (taxes plus out of pocket). Yet we found that subjects, even when told explicitly of this possibility, underadjusted the residual tax system such that progressivity fell with privatization (McCaffery and Baron 2006).

No doubt there are more biases in the popular perception of tax to be noted and verified. For example, I strongly suspect that there is a *form aversion*, under which individuals more intensely dislike a tax if they must complete paperwork in order to pay for it—one of the many features that favors payroll, sales, or value-added taxes over income-based ones. (Other factors in the United States include that the payroll tax system, the *highest tax for some 80 percent of Americans*, has been labeled a “contribution,” is nominally collected one-half from one’s employer, and does not feature a salient bottom-line amount [as the income tax with its annual form does], and more [McCaffery 1994].)

## 4.2 Normative Analysis

Scholars in all veins of law and economics, behavioral or otherwise, tend to shy away from normative analysis, believing that their task is solely descriptive. But aside from a certain naïveté in separating out descriptive from normative scholarship, the questions raised by tax and public finance seem unavoidably normative (Gamage and Shanske 2011). What “should” the citizenry choose for its tax and other fiscal policies? Competition, at least the sort of competition that obtains in private markets, discussed above, will not settle the matter. Here, in contrast to the more targeted behavioral patches of the pro-savings plans discussed above, is where behavioral approaches have much to add to thinking about tax law.

I cannot in a brief chapter (also intended to survey recent literature) set forth and defend a complete and compelling normative approach to tax. Nevertheless, I can and do note a certain “optimal” approach to law and economics, as set forth by others, most famously Kaplow and Shavell (1994, 2002), and note its potential application to tax and public finance, for the limited but important purpose of sketching out a role for a behavioral law and economics approach to tax.

In Kaplow and Shavell’s view, which builds on neoclassical welfare economics theory, the law serves two functions: allocative and distributive. In its allocative function,



the goal of the law should be to maximize wealth, or, equivalently, to obtain efficiency—to make the celebrated social pie as large as possible. Kaplow and Shavell argue that this should be the sole task of all private law subject matters, perhaps at least if important individual rights are not at stake. Then the tax system can be used to most efficiently redistribute the greater social pie resulting from step one. I have argued with my coauthor Jon Baron (McCaffery and Baron 2005) that such an approach can extend to tax as well: tax systems should be chosen for their efficiency properties, and then redistribution should follow. (See also Gamage and Shanske 2010.) The issue is that a behavioral approach to tax calls any such neat bifurcation into question.

The most simple statement of the positive findings about tax law design canvassed above (which, like almost all behavioral economics findings, finds ready support in common sense and popular understanding) is that people do not like taxes. Politicians will search for low-salient taxes to fund the high-salient spending programs that people also like (more on this anon). Now there is a widespread tendency, even in the academy, to think that if tax and other public finance systems appeal to popular perceptions, it is a good thing. There is a psychic gain from putting the pain of tax in its most pleasing light. Thus a recent rash of scholarship, building on the rather limited findings of Chetty, Looney, and Kroft (2009), has argued that governments “should” seek low-salient taxes, and take whatever steps needed to reduce their “political” salience as well (Galle 2013; Chetty, Looney, and Kroft 2009; Gamage and Shanske 2010, 2011; Schenk 2010). But any immediate practical application of the research seems wrong, or far too quick, for at least five general reasons.

*First*, psychologically pleasing taxes, including “low salient” ones, have real effects. In particular, they can be inefficient, which violates the first prong of the optimal welfare economics analysis. Chetty, Looney, and Kroft (2009) consider a rather modest sales or value-added tax. But the corporate income tax in almost all developed economies is a better, bigger, and more pervasive example of a hidden tax. Although the tax seems to please people (specifically in the sense that it does not strike them as a “tax,” or at least not one that they pay), a corporate tax has real effects on prices and other allocative decisions. If the distorting costs of the tax are higher than those of any alternative equal revenue raiser then, *ceteris paribus*, society is paying a real cost, in terms of welfare, for its psychic preferences. This is a straightforward application of the Kaplow-Shavell neoclassical welfare approach. The first prong of the optimal welfare economics approach (maximize the social pie) cannot be followed because the people will not accept efficiency-enhancing reforms; a potential *Pareto-improving* reform is not taken.

The Chetty, Looney, and Kroft (2009) and other related findings are not surprising. Much of behavioral economics suggests that individuals will underreact to, and perhaps completely ignore, a tax to be imposed later, even *minutes* later. But moving this fact out to any meaningful set of public policy recommendations is problematic at best. On one hand, it all seems rather obvious and unnecessary. Politicians have every incentive to choose low-salient taxes to begin with. This is a statement of the central *problem*, not the *cure*. Yet scholars have taken the Chetty, Looney, and Kroft findings

to argue that low-salient taxes have virtually no “substitution” effects and thus no costs whatsoever (Chetty 2009; Gamage and Shanske 2010).<sup>7</sup> Low market salience taxes seem to promise a free lunch. (The public finance concept is that taxes have two effects: a substitution one, as to which goods or services are purchased, and an income one, from taking money away from people. The income effect is a pure “transfer,” so neoclassical economics tends to focus on the substitution effect, where misallocations can arise.) But of course there is no free lunch—it is a behavioral bias, an excessive degree of optimism, which leads anyone to think that there is or can be one.

The “use low-salient taxes” recommendation is thus another illustration of the dangers of an overly quick application of behavioral insights to tax law. It is quite a leap from the idea that buyers do not include, in the short term, taxes that later get applied to the goods they have selected, even where they are given information that the tax will be applied minutes later at the cash register, to the idea that there is “no” efficiency or substitution effect in such cases. Certainly, hiding the tax can mute the substitution effect, and it is possible that there could be no substitution effect at all if the taxpayer is not told at the proper moment about the tax. People continue to buy the now taxed good as before. But there will be an effect, at least an income effect, as discussed next. In this case, we can say in response to a statement that low-market-salient taxes have no substitution effects, and therefore should be chosen by a welfare-maximizing government, that “everything follows except the ‘therefore,’” because income effects are central to the normative analysis of tax systems. Further, query whether or not any “low” or “no” market salient tax would remain so if the device were pushed out to an extreme, as a real-world political system, starved for revenue, would be motivated to do: a kind of “winner’s curse” (Thaler 1980) may haunt the tax system, whereby the taxes that best survive democratic political processes are the most flawed, from a normative perspective.

Continuing the analysis, note that, *second*, equity can suffer in a move towards low-salient taxes. Equity can be pitted against efficiency in a trade-off not mandated by an optimal approach. Even if there is no substitution effect from low market salient taxes, as the optimistic scholars argue, there *must* be an income effect, or the government is not raising the revenue that is the point of the whole exercise. This is especially true as “low” or “no” salience taxes are almost certain to be ones featuring a wide base and low rates. You can fool all of the people for a few cents here and there, but you cannot fool too many of them for too long over large sums. As noted above, individuals do not typically properly add together taxes to form consistent judgments about them. Again we see policymakers suffering from heuristics and biases themselves. When

<sup>7</sup> I do not mean this as a criticism of the particular scholars or scholarship working in this vein; these scholars and their work do reflect many of the concerns raised herein. I do however worry about the effects of a less detailed reading of the work in the public political discourse, where politicians have shown a tendency to favor hidden taxes, whatever their costs, and typically without taking counter-active corrective measures. I thank Brian Galle in particular for helpful comments on this point.

scholars are considering and arguing for low-salient taxes on the basis of their having “no substitution effect,” they are looking at small taxes on individual items, the pennies or at most cents added on at the cash register. But in moving out to discuss *income* effects, one should, rationally and fairly, consider the *aggregate* dollars lost to the tax, indeed, all taxes. A 10 percent sales tax added to some but not all goods in a grocery store may not lead an individual consumer to substitute away from taxable to nontaxable goods, as she rationally should. But a household that spends, say, \$12,000 a year on groceries subject to the levy will have \$1,200 less in its budget set over an annual period. Something has to give. There are or ought to be normative concerns.

In the abstract, psychically pleasing taxes may not, and generally will not, be as progressive as subjects themselves desire taxes to be. Chetty (2009) and others (Gamage and Shanske 2011) speculate that the income effect of low-salient taxes may not be distortionary if it results in the consumer cutting back on “luxury” items, or if the consumer is not “credit constrained.” Here again we see cognitive biases in play. Gamage and Shanske (2011) rightly applaud the empiricism of Chetty, Looney, and Kroft (2009). Yet atop a rather modest empirical finding they attempt to construct a prescriptive mountain, resting on. . . hopeful speculation. What low- or middle-income consumer today, in any advanced democracy, already subject to a panoply of low-salient taxes, is spending a great deal of residual income on “luxuries” or is not “credit constrained”? Are these not the very same consumers and individuals, by and large, who are saving “too little” by their own lights, as the *other* principal behavioral economics-inspired policy recommendation for tax law, that of tax-favored savings accounts, considered? But if people are saving too little, they are credit-constrained or at least have a nonluxury item, savings, to which to devote their resources. Yet the policy recommendation for low-salient taxes is not to use the money “costlessly” appropriated from ordinary consumers to fund savings plans for these very same consumers. This concept is in fact intriguing, but one that could suffer from arbitrage and other problems (e.g., government appropriation). The policy reason behind the proposal is, rather, to use the tax revenue for general governmental needs in an age of massive fiscal deficits.

Gamage and Shanske (2010) argue, in the manner of Kaplow and Shavell (1994), that any skew in distribution from the imposition of low or no market salient taxes could be counterbalanced by changes to the residual tax system, such as the income tax. They do note a caution that “even when offsetting tax-rate adjustments are theoretically capable of resolving distributional concerns, political or administrative limitations may prevent the implementation of the offsetting tax-rate adjustments” (Gamage and Shanske 2010, 52). Indeed, but reality is worse, and less hopeful than that quick statement would have it be. Precisely as persistent myopia, given the status quo, can undermine behavioral patches designed to address myopia, so could and almost certainly would *salience* undercut any attempt to offset low-salient, broad-based, low rate level regressive taxes with other taxes. Those “other” taxes will be, by necessity, highly salient and concentrated ones. This is precisely the kind of tax increase that the proponents of the proposal themselves would also find to be widely unpopular.

If the isolation or disaggregation effect were not so widespread, any equity effect from the widespread imposition of low-salient taxes may not matter all that much. Society could have as many regressive taxes or surcharges as it desired, as long as it had a single system, such as the individual income tax, in which to redistribute. But we have seen many times over that ordinary subjects will have a hard time understanding extreme progressivity in any single system, when viewed in isolation (McCaffery and Baron 2003, 2006). The reformer concerned with redistribution needs to look at all tax systems, individually, because the polity will not adequately integrate them. In the case of these low-salient tax policy proposals, the extent of the increase to the individual income tax needed to offset the lack of progressivity in any significant low-salient tax would have to be extreme—and hence, on this account, highly salient. If citizens do not support highly salient taxes, how can highly salient taxes be used to correct for low-salient ones? Recent political history in the United States, at least, for well over two decades now, suggests that it is virtually politically impossible to raise marginal tax rates under the individual income tax, the most salient, and progressive, major tax in existence (McCaffery and Hines 2010).

The same tension is evident in what has been called the *privatization effect* (McCaffery and Baron 2006). The two-part optimal welfare economics analysis suggests that efficiency, alone, should dictate whether or not government provides any particular good or service. However, because ordinary subjects once again have a difficult time integrating the effect of spending cuts, or government downsizing, on the residual tax system, bottom-line redistribution can suffer on account of even an efficiency-enhancing reform. The Paretian constraint will not hold, the rich will get richer, and the poor, poorer. This is troubling.

*Third*, the resolution of public finance matters can be fragile, and volatile, as equivalent *frames* can shift public opinion. Instability in public finance systems is itself, *ceteris paribus*, a bad—a welfare-reducing phenomenon. Note that psychological studies suggest that preference shifts or reversals can obtain with no change in the underlying substance. It is not a matter of people seeing the light and adopting “better” resolutions of public finance systems. People will simply choose more progressivity if they can be led to think in percentage terms, and less in dollar terms; they will choose policies that can be understood as bonuses and reject them when they see them as penalties. This back and forth, on purely formal grounds, is problematic. Worse, it leads to another concern.

*Fourth*, given the importance of framing and related effects, politics will reward rhetoric over substance. “Great communicators” will be prized, not because they advocate “better” policies, but because they make their policies *sound better* to voters. In and of itself, this diverts political resources from the potentially welfare-enhancing study of substantive policy effects, to the purely formal rhetorical presentation of matters. And this leads, lastly, to an especially great concern.

*Fifth*, finally, and perhaps most disturbingly: a skilled politician or political party can manipulate public opinion and get a public finance system in place that is in conflict both with the public good and with prevalent democratic preferences. Suppose

that some politician or party wanted to reverse the course of a particular society and reduce the degree of redistribution prevailing throughout a public finance system. Insights from behavioral economics provide an eerie road map for success. Many findings suggest that a policy position to lessen social redistribution would likely lose in a straight up or down vote, both because the majority of people favor at least moderate redistribution and because of a persistent status quo bias. But the rhetorically skilled politician could influence a collective preference reversal. She might first choose hidden or low-salient taxes, with a regressive incidence, and raise money through a series of relatively flat surcharges that are not labeled as taxes. People would support these, and a surplus might even result. Larger surpluses might follow from selective “privatization” of current publicly provided goods and services, reducing the need for taxes. Cuts could then be made to the most salient tax alone, the individual income tax, and that tax could be brought to reflect moderate progressivity, even as its importance in the overall budget declined. Indeed, the politician could take this all a step further, and separate out the topics of tax and spending cuts, cutting taxes (again, the highly salient income tax) now, and postponing spending cuts until later. The resulting deficit, itself a form of a hidden tax, would curtail government growth, and could lead to replacement taxes that are less progressive (and salient) than the initial baseline. And so on: we would wake up one day with a smaller government, less dependent on the single remaining progressive tax system, and that tax system would continue to have only moderate levels of progressivity while various replacement taxes were more regressive. Over all, this series of steps would lead to dramatically less redistribution than the people themselves wanted, at the outset, and the cumulative changes would also fail to meet the basic Paretian constraint.

Of course, the astute observer might notice that this is what has been done in the United States, beginning with Ronald Reagan in 1981.

### 4.3 A Case Study

I shall use a more recent and specific example to illustrate some of the points just articulated.<sup>8</sup>

As the United States careened toward a self-created “fiscal cliff” set to obtain on January 1, 2013, the press was full of febrile leaks and rumors of impending deals between the Obama administration and the Democrats, on one side, and the Speaker of the House, John Boehner, and the Republicans on the other side. Both sides faced a problem, and a strictly rhetorical one: how to “score” the tax changes being considered.

A large reason behind the fiscal cliff “crisis” was that the tax cuts of 2001 and 2003, ushered in during the administration of President George W. Bush, were each set to expire on January 1, 2013, meaning that there would be a return to the higher tax rate

<sup>8</sup> This subsection draws from McCaffery (in progress).

levels that obtained in the year 2000. Both political parties clearly and consistently agreed that no taxpayer earning less than \$250,000 should see a tax increase. There was therefore no real chance that what Congress would ultimately enact would be anything other than a tax *cut*, on balance; elected officials were not going to let the tax cuts expire for the masses. Since, as the law stood, all income taxpayers were set to face a tax *increase*, and since all political roads were leading to a resolution in which only *some* of them would get the tax increase that was due (that is, the “rich,” or “high-income”), the new law cutting the taxes of most while leaving in place the already legislated tax increases of a few would have to be considered as a tax *decrease*. This is indeed how the Congressional Budget Office officially “scored” the ultimate law. Only that would not do, politically.

So both sides, happily aided and abetted by the media, began talking about what they were doing in terms of making changes from the 2012 baseline—2012 being a year in which both the 2001 and 2003 tax cuts were in full force and effect. Consequently, this conceit stood the legal reality on its head perfectly: it scored the tax changes to be enacted as if the Bush era tax cuts were *permanent*, as opposed to what they were, which was *expiring*. But, more importantly, it would allow both sides to claim a tax *increase* on the rich from the failure to extend their tax breaks, while not highlighting the sound analytic fact that everyone else was getting a tax cut.

This might just be another story of business as usual inside the Beltway, not playing a major role in this chapter, except for the twists and turns occasioned by one other further inconvenient truth. Taxes on January 1, 2013, were *also* set to go up by virtue of the expiration of the “payroll tax holiday” that President Obama had enacted for 2011, and which Congress had extended (in late 2011) for 2012. The “holiday” was a 2 percent break (in absolute terms) from the 6.2 percent employee share of the Social Security payroll tax, applicable to earnings (in 2012) up to approximately \$110,000 per individual. The “holiday” could thus save an individual over \$2,000 and \$4,000 for a couple, each member of whom earned \$100,000. This provision was set to expire just like the Bush tax cuts. Only both sides, Democrats and Republicans, very quietly agreed to let it die a quiet death.

The resolution of the rhetorical embarrassment, occasioned by the fact that what Congress was doing was actually a tax decrease in a time of fiscal crisis, was to use a 2012 baseline to measure increases off of it. This same accounting construct (measuring changes from 2012’s law) would allow President Obama to count the tax increases from the payroll tax holiday’s expiration as tax increases. This would seem to be a *good* thing, as politicians were scrambling to meet proposed revenue-raising targets. More specifically, President Obama led off the negotiations with a proposed revenue-raising target of \$1.6 trillion over ten years—all tax increases get scored over a ten-year period, these days. Republicans had countered with \$800 billion, and both sides seemed to be inching towards the obvious and inevitable compromise figure of \$1.2 trillion.

The optics of meaningful aggregate tax increases and deficit reduction, not to mention logical consistency, suggested “counting” the payroll tax increase. But the optics of redistribution suggested otherwise; in order to maintain a rhetorical claim about



the “top 2 percent” bearing all of the new burdens. The truth was going to be what it was: the payroll tax holiday was going to expire, but the (misleading) optics of redistribution prevailed. The Obama administration did not “count” the tax increases from the payroll tax holiday’s expiration as part of the fiscal cliff tax increases, although this meant, logically, the numbers being used could only be described as “tax increases over a 2012 baseline, but not including matters not addressed in a new law,” or something similar. The Republicans, hardly chafing at the bit to tout any tax increase, readily played along.

What are the dollars and cents of all of this? Official sources and the mainstream media reported tax increases from the law ranging from \$600 to \$700 billion (a sum total below even the Republicans’ opening bid), almost all falling on the “rich” in some sense. Here is an illustrative description, from the progressive group, Center for American Progress (2013):

The American Taxpayer Relief Act of 2012—the fiscal cliff legislation agreed to in a deal between President Obama and Senate Minority Leader Mitch McConnell (R-Kentucky)—will raise approximately \$617 billion in higher revenues from 2013 to 2022, compared to what the tax code would have generated if we had simply extended all the Bush tax cuts, which were scheduled to expire at the end of 2012. *More than 90 percent of the increase will come from households making at least \$1 million a year.* (Emphasis supplied)

As for the expiration of the payroll tax holiday? This was estimated to bring in an additional \$100 billion per year or \$1 trillion over a decade, meaning President Obama could have claimed to make his \$1.6 trillion mark after all. But he did not, because only a very small percentage of the \$1 trillion gained from letting the payroll tax cut expire would come from households making at least \$1 million a year.

In sum, roughly 62.5 percent of the aggregate tax increase in the fiscal cliff package came from a tax that applies only to wage earnings and which has a floor of zero (meaning it applies to the first dollar earned, regardless of family size or anything else) and a ceiling of \$110,000. But this major tax increase was not listed in the official reports or scoring, in order to maintain a rhetorical claim that 90 percent of the tax increases would fall on millionaires only. We end up with not so much a hidden tax, although the payroll tax is much better hidden than the income tax in multiple regards, but rather a hidden tax *increase*, buried in the self-serving rhetoric of both parties inside the Beltway. This is what politicians do with their behavioral approach to tax law design.

## 5 FISCAL BALANCE

---

The final application, or potential application, of a behavioral economics approach to tax law concerns the wider question of fiscal balance, now arguably the major issue facing advanced capitalist democracies worldwide.



In some sense this is simply an instance of the prior topic (tax law design) because a failure to generate adequate revenues is a primary cause of the sovereign debt crises brewing globally (and this view of the problem has been explicitly invoked in defense of the low-salient tax policy recommendation (Schenck 2010)). But the comprehensive importance of the subject matter warrants separate attention, and also brings in the other side of the fiscal equation: government spending programs. Behavioral economics, in its canonical findings such as *loss aversion* and the *status quo bias*, are clearly relevant to this piece of the puzzle. See Zamir (2012) and chapter 11 in this volume, discussing inter alia loss aversion and its relevance to the “tax expenditure” budget—tax expenditures being a form of government spending through tax breaks, itself both a widespread phenomenon and one rife with behavioral elements.

Various polls and other experimental findings support the idea that most individuals, out of context and in the abstract, support balanced budgets (Baron and McCaffery 2008). Yet we do not make decisions in the abstract, and policies do not take effect out of context. How can deficits get started in a democratic citizenry that opposes them? Behavioral economics supplies a set of answers. The key is to match specific tax cuts today, which individuals will support out of tax aversion, with the abstract, general idea of spending cuts tomorrow, which individuals will also support today. On one hand, if tax cuts *today* must be matched by specific spending cuts *today*, then an opposition to both specific spending cuts and to deficits is likely to preserve the status quo of balance. On the other hand, if the tax and spending decisions can be separated in time and (logical) space, then the specificity of the spending cuts can recede, and a *disaggregation bias* effect can take hold. Individuals tend to focus on the tax cuts alone, where a generic tax aversion will lead them to support cuts, resulting in a budget deficit. Once this deficit is created, the preference for fiscal prudence causes people to want to raise taxes and limit spending. But these desires are not strong enough to reduce the deficit to zero, even when people are asked about the “long run.” And a more persistent and pervasive status quo bias means that it will be difficult to ever obtain fiscal balance once deficits have become accepted.

The classic element of a behavioral approach, inconsistency, is present here. The inconsistency does not seem to follow from a simple optimism bias. It is not that people have a naive belief that things will be better tomorrow, and deficits will miraculously be closed without the pain of tax increases or spending cuts. Rather, it is a failure to properly anticipate the depth of the difficulty in making specific cuts tomorrow—the depth, that is, of an *endowment effect*. At a high level of generality, deficits arise or increase when specific (salient) tax cuts go along with abstract (nonsalient) sets of spending cuts. This conceptualization suggests two broad ways for governments to avoid deficits. One is to keep everything abstract: to pass laws in the form of constitutional restrictions about balanced budgets, a form of collective self-binding to the mast (Elster 1984). Many state governments in the United States are indeed required to have balanced budgets each year, and the US government has occasionally tried to bind itself in advance by various budgetary rules. This approach could also have the virtue of making a nonelected body (the judiciary) the “bad cops,” playing the role of

an arbitrage mechanism to help debias the citizenry away from its pervasive biases. Central bankers around the world, or collective entities such as the European Union or International Monetary Fund (IMF), might play this role, too, imposing “austerity conditions” on member states, or as a condition of financing. These tactics have had mixed political and financial success to date, but also point towards a depressing fact—democracy itself might have to be curtailed because of the widespread and harmful effects of behavioral biases on fiscal systems.

An alternative takes the opposite approach: to make everything concrete and specific. We could break taxes down into categories earmarked for particular services, such as, nominally, in the case of the various wage taxes in the United States (which are earmarked for Social Security payments and the like). If citizens come to think of each tax as linked with a particular service, they may be less willing to cut taxes; indeed, this could explain why the Social Security and Medicare tax in the United States, now the largest tax for most American taxpayers, is the one major federal tax that had never been cut until just two years ago. This alternative would probably lead to a larger, more active government than the first method (binding in the abstract).

These are complex and difficult topics, which point to ultimately complex and difficult political solutions or cures. But these are also ideas worth exploring, in the ongoing quest for a practical, real-world and normative role for a behavioral economics approach to tax law. It is past time to accelerate the pace of our thinking about behavioral economics and tax.

## ACKNOWLEDGMENTS

---

An earlier version of this chapter was presented at a workshop at Loyola Law School; I thank Katie Pratt, Ted Seto, David DePianto, and other participants and also Brian Galle for helpful comments; the editors of the volume for many helpful suggestions; and Stephan Airapetian, Stephanie Myer, and June Yang for excellent research assistance.

## REFERENCES

---

- Alm, J., J. M. Vazquez, and B. Togler, eds. 2010. *Developing Alternative Frameworks for Explaining Tax Compliance*. New York: Routledge.
- Bar-Gill, O., and O. Board. 2012. Product-use information and the limits of voluntary disclosure. *American Law and Economics Review* 14:235–70.
- Baron, J., and E. McCaffery. 2006. Masking redistribution (or its absence). *Behavioral Public Finance*. New York: Russell Sage Foundation, 85–12.
- Baron, J., and E. McCaffery. 2008. Starving the beast: The political psychology of budget deficits. *Fiscal Challenges: An Inter-disciplinary Approach to Budget Policy*. Cambridge, MA: MIT Press, 221–39.

- Barro, R. 1974. Are government bonds net wealth? *Journal of Political Economy* 82(6): 1095–117.
- Becker, G. S. 1962. Irrational behavior and economic theory. *Journal of Political Economy* 70:1–13.
- Bell, E., A. Carasso, and C. E. Steurle. 2004. Retirement savings incentives and personal savings. *Tax Notes* 105:1689–89.
- Camerer, C. 1998. Prospect theory in the wild: Evidence from the field. *California Institute of Technology, Division of Humanities and Social Sciences* 5:1–13.
- Camerer, C., S. Issacharoff, G. Loewenstein, T. O’Donoghue, and M. Rabin. 2003. Regulation for conservatives: Behavioral economics and the case for “asymmetric paternalism”. *University of Pennsylvania Law Review* 151:1211–54.
- Chetty, R. 2009. The simple economics of salience and taxation. NBER Working Paper No. 15246.
- Chetty, R., A. Looney, and K. Kroft. 2009. Salience and taxation: Theory and evidence. *American Economic Review* 99:1145–77.
- Choi, J., D. Laibson, B. Madrian, and A. Metrick. 2002. Defined contribution pensions: Plan rules, participant choices and the path of least resistance. *Tax Policy and the Economy* 16:67–114.
- Elster, J. 1984. *Ulysses and the Sirens: Studies in Rationality and Irrationality*. New York: Cambridge University Press.
- Feldman, Y., and D. Teichman. 2008. Are all “legal dollars” created equal? *Northwestern University Law Review* 102:1–32.
- Frey, B. S., and B. Torgler. 2007. Tax morale and conditional cooperation. *Journal of Comparative Economics* 35:136–59.
- Galle, B. D. 2013. Carrots, sticks and salience. Boston College Law School Legal Studies Research Paper No. 290.
- Gamage, D., and D. Shanske. 2010. The case for reducing the market salience of taxation. *National Tax Association Proceedings* 103:48–53.
- Gamage, D., and D. Shanske. 2011. Three essays on tax salience: Market salience and political salience. *Tax Law Review* 65:19–98.
- Gruber, J., and B. Köszegi. 2004. Tax incidence when individuals are time-inconsistent: The case of cigarette excise taxes. *Journal of Public Economics* 88:1959–87.
- Guttel, E., and A. Harel. 2008. Uncertainty revisited: Legal prediction and legal postdiction. Hebrew University of Jerusalem.
- Hill, C. A. 2010. What cognitive psychologists should find interesting about tax. *Psychonomic Bulletin & Review* 17:180–85.
- Hoffman, E., and M. L. Spitzer. 1982. The Coase Theorem: Some experimental tests. *Journal of Law and Economics* 25:73–98.
- Isoni, A., G. Looms, and R. Sugden. 2011. The willingness to pay–willingness to accept gap, the “endowment effect,” subject misconceptions, and experimental procedures for eliciting valuations: Comment. *American Economic Review* 101:991–1011.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1990. Experimental tests of the endowment effect and the Coase Theorem. *Journal of Political Economy* 98:1325–48.
- Kahneman, D., D. Schkade, and C. Sunstein. 1998. Assessing punitive damages. *Yale Law Review* 107:2071–152.
- Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47:263–91.

- Kahneman, D., and A. Tversky, eds. 2000. *Choices, Values and Frames*. New York: Cambridge University Press and Russell Sage Foundation.
- Kaplow, L., and S. Shavell. 1994. Why the legal system is less efficient than the income tax in redistributing income. *Journal of Legal Studies* 23:667–81.
- Kaplow, L., and S. Shavell. 2002. *Fairness versus Welfare*. Cambridge, MA: Harvard University Press.
- Knetsch, J. L., and W. Wong. 2009. The endowment effect and the reference state: Evidence and manipulations. *Journal of Economic Behavior and Organization* 71:407–13.
- Koh, W., and W. Wong. 2011. The endowment effect and the willingness to accept–willingness to pay gap: Subject misconceptions or reference dependence? Working paper, Department of Economics, National University of Singapore.
- Konrad, K. A., and S. Qari. 2012. The last refuge of a scoundrel: Patriotism and tax compliance. *Economica* 79:516–33.
- Korobkin, R. 2000. Behavioral economics, contract formation, and contract law. *Behavioral Law and Economics*, C. Sunstein, ed. Cambridge: Cambridge University Press, 116–43.
- Jolls, C., C. R. Sunstein, and R. H. Thaler. 1998. A behavioral approach to law and economics. *Stanford Law Review* 50:1471–550.
- Liebman J., and R. Zeckhauser. 2004. Schmeduling. Harvard University and Nation Bureau of Economic Research.
- Loewenstein, G., and J. Elster, eds. 1992. *Choice over Time*. New York: Russell Sage Foundation.
- McCaffery, E. J. 1994. The UCLA tax policy conference: Cognitive theory and tax. *UCLA Law Review* 41:1861–947.
- McCaffery, E. J. 2006. Behavioral economics and fundamental tax reform. *Fundamental Tax Reform: Issue, Choices, and Implications*, J. W. Diamond and G. R. Zodrow, eds. Cambridge, MA: MIT Press, 455–84.
- McCaffery, E. J. In progress. *Taxing Wealth Seriously*.
- McCaffery, E. J., and J. Baron. 2003. The Humpty-Dumpty blues: Disaggregation bias in the evaluation of tax systems. *Organizational Behavior and Human Decision Processes* 91:230–42.
- McCaffery, E. J., and J. Baron. 2004. Framing and taxation: Evaluation of tax policies involving household composition. *Journal of Economic Psychology* 25:679–705.
- McCaffery, E. J., and J. Baron. 2005. The political psychology of redistribution. *UCLA Law Review* 52:1745–92.
- McCaffery, E. J., and J. Baron. 2006. Thinking about tax. *Psychology, Public Policy and Law* 12:106–35.
- McCaffery, E. J., and J. Hines. 2010. The last best hope for progressivity in tax. USC Law Legal Studies Paper No. 09-8.
- McCaffery, E. J., D. Kahneman, and M. L. Spitzer. 1995. Framing the jury: Cognitive perspectives on pain and suffering awards. *Virginia Law Review* 81:1341–420.
- McCaffery, E. J. and J. Slemrod, eds. 2006. *Behavioral Public Finance*. New York: Russell Sage Foundation.
- Mirrlees, J. A. 1971. An exploration in the theory of optimum income taxation. *Review of Economic Studies* 38:175–208.
- Modigliani, F. 1966. The life-cycle hypothesis of saving, the demand for wealth and the supply of capital. *Social Research* 32(2):161–217.
- Noll, R. G. and J. E. Krier. 1990. Some implications of cognitive psychology for risk regulation. *The Journal of Legal Studies* 19:747.

- Pigou, A. C. 1920. *The Economics of Welfare*. London: Macmillan.
- Posner, R. A. 1972. *An Economic Analysis of Law*. Boston: Little, Brown.
- Popper, K. R. 1971. *The open society and its enemies*. Princeton, NJ: Princeton University Press.
- Ramsey, F. P. 1927. A contribution to the theory of taxation. *Economic Journal* 37:47–61.
- Robben, H. S. J., et al. 1990. Decision frame and opportunity as determinants of tax cheating: An international experimental study. *Journal of Economic Psychology* 11:341–64.
- Schelling, T. C. 1981. Economic reasoning and the ethics of policy. *Choice and Consequence*. Cambridge, MA: Harvard University Press, 1–26.
- Schenk, D. 2010. Exploiting the salience bias in designing taxes. New York University Law and Economics Working Papers.
- Scotchmer, S., and J. B. Slemrod. 1988. Randomness in tax enforcement. NBER Working Paper No. w2512.
- Sen, A. K. 1977. Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs* 6:317–44.
- Shefrin, H., and R. H. Thaler. 2007. The behavioral life-cycle hypothesis. *Economic Inquiry* 26:609–43.
- Slemrod, J. 2007. Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives* 21(1):25–48.
- Smith, A. [1776] 2009. *The Wealth of Nations*. Blacksburg, VA: Thrifty Books.
- Sunstein, C., ed. 2000. *Behavioral Law and Economics*. Cambridge: Cambridge University Press.
- Sunstein, C., and R. H. Thaler. 2003. Libertarian paternalism is not an oxymoron. *University of Chicago Law Review* 70:1159–202.
- Thaler, R. H. 1980. Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization* 1:39–60.
- Thaler, R. H. 2008. Mental accounting and consumer choice. *Marketing Science* 27:15–25.
- Thaler, R. H., and S. Benartzi. 2004. Save more tomorrow: Using behavioral economics to increase employee saving. *Behavioral Economics* 112:164–87.
- Tversky, A., and D. J. Koehler. 1994. Support theory: A nonextensional representation of subjective probability. *Psychological Review* 101:547–67.
- Zamir, E. 2012. Loss Aversion and the Law. *Vanderbilt Law Review* 65:829–94.
- Zeiler, K., and C. R. Plott. 2011. The willingness to pay–willingness to accept gap, the “endowment effect,” subject misconceptions and experimental procedures for eliciting valuations: Reply. *American Economic Review* 101(2):1012–28.

## CHAPTER 24

---

# LITIGATION AND SETTLEMENT

---

JENNIFER K. ROBBENNOLT

## 1 INTRODUCTION

---

WHEN and how legal cases are brought, invested in, settled, and taken to trial has been a frequent topic of much social scientific analysis—with contributions from economics (see, e.g., Bebchuk 1984), sociology (see, e.g., Felstiner, Abel, and Sarat 1981; Merry and Silbey 1984), political science (see, e.g., Boyd and Hoffman, 2013), and psychology (see, e.g., Guthrie 2000, Rachlinski 1996; Saks 1992).

Litigants and their attorneys must decide whether and when to file suit; whether and when to file motions and how to respond to motions filed by the other side; what discovery and trial preparation to do; what offers or demands to make; and whether to accept a proposed settlement or proceed toward trial. Along the way many disputes are resolved in some way. Most civil disputes do not result in lawsuits (see Miller and Sarat 1981; Hensler et al. 1991). And most cases that are filed as lawsuits do not go to trial. Most cases are, for example, dropped, resolved by motions, or settled in some way (see Eisenberg and Lanvers 2009; Galanter 2004; Langbein 2012). Indeed, the rate of trial is declining (Galanter 2004). For those cases that do result in trial verdicts, litigants must also decide whether and on what grounds to appeal or whether to propose or accept a posttrial settlement.

Among those cases that go to trial, some cases fail to settle even though it would have been financially beneficial to the parties to have settled. Indeed, recent work has compared verdicts with final settlement offers that were rejected—finding that both plaintiffs and defendants frequently reject settlement proposals that would have been financially advantageous (Kiser, Asher, and McShane 2008). Among those cases that settle, some are settled on terms that are favorable to both sides. In other instances, valid claims are not pursued or are settled on terms that are disadvantageous to one or both sides.

Standard economic models of litigation and settlement conceive of litigation and settlement decision-making as a process in which legal actors—litigants and their

attorneys—rationally maximize expected utility. While such models provide a helpful structure for understanding litigation decision-making, litigants and their attorneys may not always behave and decide in ways that follow the assumptions of the economic models. As we will see, the decisions that are made as a case progresses through the legal system are more complex than the standard models provide. Predicting expected outcomes, setting goals, interacting with adversaries, gathering and using information, and weighing options all present challenges for decision-makers. In addition, the dynamic interaction between lawyer and client can influence whether and how cases are settled.

## 2 STANDARD ECONOMIC MODELS OF LITIGATION AND SETTLEMENT

---

Standard economic models of litigation and settlement (see Cooter and Ulen 2012 for a review) posit that litigants engage in expected value calculations when making decisions. Claiming and decisions about whether to defend cases are influenced by litigants' assessments of the expected value of the claim and the costs of filing a complaint or responding to the suit. Similarly, decisions about whether to settle a case or to continue forward toward trial are influenced by litigants' predictions about the possible outcomes of trial, the value of the settlement under consideration, and the relevant costs. Thus, as a simplified example, a plaintiff would calculate the expected value of trial as follows:

$$EV = (\text{probability of win} \times \text{expected outcome}) + (\text{probability of loss} \times \$0) - \text{costs.}$$

A defendant's calculation would look similar—the costs of going to trial would be calculated as

$$EV = (\text{probability of win} \times \$0) - (\text{probability of win} \times \text{expected size of verdict}) - \text{costs.}$$

For each litigant, the risk discounted value of trial would be compared to the value of a proposed settlement and the litigant would choose the option (trial or settlement) with the highest expected value.

Under the economic analysis, high rates of settlement are driven, at least in part, by the costs associated with each stage of litigation. These costs include not only the financial costs of bringing or defending the suit (e.g., legal fees, discovery costs, experts), but also any implications for future cases, any expected reputational costs, the time and effort allocated to the lawsuit, and the unpleasantness of the process itself (see, e.g., Galanter 1986; Hannaford-Agor and Waters 2013; Spier 2007; Trubeck et al. 1983). To see how costs might drive the possibility of settlement, it is important to understand that the range within which bargaining can occur is defined by the difference between the litigants' subjective expected values of trial. That is, settlement is possible in the range that falls between the lowest amount for which the plaintiff will settle and the



highest amount for which the defendant will settle. This range is variously known as the *bargaining range*, the *zone of possible agreement* (ZOPA), or the *surplus*. Because the costs of trial will increase the amount a defendant is willing to pay and decrease the amount a plaintiff will insist upon receiving, the costs of trial tend to increase this bargaining range and make settlement more possible. Indeed, litigation costs saved add to the surplus and constitute joint gain.

Standard economic analysis posits two general problems that can lead to errors in settlement decision-making. First, the parties are likely to have imperfect and asymmetric information relevant to their decision—information about behavior, injuries, witnesses, and so on (see Bebchuk 1984; Spier 2007). Divergent information is likely to lead to differences in predictions about likely outcomes. Second, even if information were complete and predictions about litigated outcomes were perfect, strategic bargaining decisions could still drive suboptimal settlement decision-making. In particular, failure to settle may result from difficulties in dividing the surplus. Strategies such as refusing to consider settlement, threats, misrepresentation, delay, or simply holding out for more may be used by the parties as they attempt to claim a greater share of even a well-defined surplus (see Cooter, Marks, and Mnookin 1982; Gross and Syverud 1991).

The story told by the economic model of litigation bargaining—a story in which actors make rational decisions to maximize the expected values of their outcomes—provides a helpful picture of the litigation decision-making context. And it is true that many litigation “mistakes” are the result of incomplete information and strategic bargaining decisions. Recognizing, however, that there are a variety of behavioral or psychological (see Rachlinski 2011) phenomena that make it challenging for litigants and their attorneys to identify and choose the best litigation options can make for a richer and more accurate understanding of litigation decision-making.

### 3 POSITIVE ILLUSIONS AND PREDICTIONS

---

Predictions about what will happen at trial or on appeal—and predictions about the outcomes of motions and other interim decisions—are central features of settlement decision-making. Such predictions must be made in order to compare the expected value of trial, and settlement must take into account anticipated legal costs, relevant law, and case facts (Cooter and Rubinfeld 1989). One relatively straightforward reason why cases might proceed to trial despite the high costs of continuing litigation is that the two sides have made different predictions about what will happen at trial—divergent predictions that either shrink or eliminate the bargaining range. Inaccurate predictions can result from the lack of information that results from poor preparation, lack of access to the relevant facts or legal rules, strategic behavior, or poor communication among the parties (see Sternlight 1999).

Behavioral scientists have also identified several psychological reasons why litigants' predictions might diverge. First, people experience a *self-serving bias* such that they perceive that which is in their own interests as being fair and expect that objective observers will favor their perspective. In the litigation setting, for example, litigants and their attorneys tend to make biased predictions about how a neutral judge would decide their cases—with plaintiffs predicting higher awards and defendants predicting significantly lower awards. Similarly, the opposing sides tend to make biased estimates of the fair settlement values of their cases (see Babcock et al. 1995a; Loewenstein et al. 1993; Babcock and Loewenstein 1997 [reviewing studies]; see also Thompson and Loewenstein 1992; Babcock et al. 1995b). Litigants are better able to remember evidence that favors their case and more likely to believe that a judge will credit their arguments (Loewenstein et al. 1993). These effects are robust enough that even observers who merely adopt the perspective of one side make these sorts of biased predictions (Thompson 1995). Importantly, the bigger the gap in perceptions, the less likely a case is to settle (Loewenstein et al. 1993).

There is also evidence that predictions about case outcomes are influenced by how narrowly defined are the outcomes to be predicted—a problem of *subadditivity*. That is, attorneys' predictions about the likelihood of a particular case outcome (for example, the likelihood of a nontrial outcome) tend to be smaller than the sum of the attorneys' estimates of the likelihood of each of that outcome's component outcomes (for example, the possibility of dismissal + settlement + case withdrawn + defendant found immune from suit). "Hence, separate evaluation of more specific possible [case] outcomes yields higher total judged probability" (Fox and Birke 2002, p. 165; see Rottenstreich and Tversky 1997; Tversky and Koehler 1994). In one study, researchers asked attorneys to predict the likelihood that a verdict would fall in one of the following ranges: up to \$25,000, \$25,000 to \$50,000, \$50,000 to \$100,000, and over \$100,000. The sum of the mean estimates for the four groups was 170% (Fox and Birke 2002). Such malleability in the substance of predictions can influence the existence or size of the bargaining range.

To make matters even more complicated, people tend to be *overly optimistic* as well as *overconfident* in their predictions (see Moore and Healy 2008). When perspective is focused on one side of a case (as is the case in litigation), parties are *less* accurate in their predictions, but hold those expectations with *more* confidence, even when the other side's arguments are easy to predict and take into account (Brenner, Koehler, and Tversky 1996). In one recent study, attorneys were shown to be overconfident in their predictions about the outcomes of their own pending cases—overstating the probability that they would meet or exceed their own minimum goals for the case (Goodman-Delahunty et al. 2010; see also Loftus and Wagenaar 1988). This inaccuracy is difficult to correct, however. These same attorneys failed to realize that their predictions were overly rosy—that is, in hindsight, the attorneys reported that they *had* met their goals (Goodman-Delahunty et al. 2010).

## 4 SETTLEMENT IN THE SHADOW OF TRIALS, SETTLEMENTS, AND THE MEDIA

---

As we have seen, the expected value calculations contemplated by the standard economic model are based in large part on parties' expectations about what will happen at trial. Thus, it is often said that litigation bargaining happens in the "shadow of the law" (Mnookin and Kornhauser 1979). In an idealized world, expectations about what is likely to happen at trial would be informed by a systematic assessment of the distribution of trial outcomes in cases with similar facts. But such expectations are also informed (and distorted) by other information to which parties are exposed.

First, because people tend to judge probabilities with reference to the ease with which they can bring to mind similar examples, possible outcomes that are more *available* in memory or more fluently brought to mind are judged to be more likely (Tversky and Kahneman 1973). Second, judgments can be influenced by *anchors*—available numbers that provide a salient (even if not relevant) benchmark for the judgment. Adjustments away from the anchor are often insufficient, resulting in judgments biased in the direction of the anchor (Tversky and Kahneman 1982).

In the negotiation context, outcomes can be influenced by anchors that are provided by initial offers or demands (Galinsky and Mussweiler 2001; Guthrie and Orr 2006; Korobkin and Guthrie 1994a; Korobkin and Guthrie 1994b; Schwab and Heise 2011), damage caps (Babcock and Pogarsky 1999), negotiator aspirations (Korobkin 2002), media reports of other cases (Bailis and MacCoun 1996; Robbennolt and Studebaker 2003), and information about private settlements (Depoorter 2010). There is evidence that both media depictions of trial outcomes and lawyer communications about private settlements provide skewed information about case outcomes, painting a picture of frequent plaintiff wins and large awards (see Ballis and MacCoun 1996; Depoorter 2010). This selective availability of information about case outcomes can distort predictions. Consistent with this expectation, even experienced attorneys tend to make skewed estimates of plaintiff win rates and the size of jury awards (see, e.g., Songer 1988).

## 5 PROSPECT THEORY AND FRAMING

---

In the context of litigation and settlement, trial outcomes are uncertain, while agreeing to a settlement provides a certain outcome. The standard economic model of bargaining assumes that litigants are either *risk-neutral* or *risk-averse* in the face of uncertainty (Cooter and Rubinfeld 1989). It turns out, however, that people's risk preferences vary

depending on whether they are facing a gain or a loss in comparison to a reference point. When it comes to a decision between options that present moderate to high probability *gains*, people tend to be *risk-averse*. But, when facing moderate to high probability *losses*, people tend to be *risk-seeking*. In addition, losses tend to have more impact than do gains of the same magnitude—this is known as *loss aversion* (Kahneman and Tversky 1979; Tversky and Kahneman 1981; Kahneman and Tversky 1984).

In litigation, the plaintiff and defendant sides face different “natural” frames. Plaintiffs face the prospect of a certain gain (if they settle) or an uncertain gain (if they go to trial). In the domain of moderate- to high-probability gains, plaintiffs are likely to be risk-averse and, therefore, likely to prefer settlement. In contrast, defendants face the prospect of a certain loss (if they settle) or an uncertain loss (if they go to trial). In the domain of moderate to high probability losses, defendants would be expected to be risk-seeking and, therefore, to prefer trial (Rachlinski 1996; see also Babcock et al. 1995; Gross and Syverud 1991; Gross and Syverud 1996). Indeed, experimental studies have found that plaintiffs choosing between gains tend to be more likely to accept a risk-neutral settlement offer (i.e., a settlement offer with the same expected value, e.g., \$50,000, as the expected value of trial, e.g., a 50% chance of winning \$100,000) than are defendants who are choosing between losses (Rachlinski 1996; see also Hogarth 1989; van Koppen 1990). Lawyers (and judges), too, appear to be susceptible to such framing effects (Rachlinski and Wistrich 2013; Guthrie, Rachlinski, and Wistrich 2001, 2009). These sorts of framing effects mean that “[r]isk seeking defendants may reject settlement proposals which they would be better off accepting. . . [Plaintiffs] may accept settlement offers that are too deeply discounted in an effort to minimize risk. . . . [L]oss aversion suggests that the settlement discouraging effect of framing on defendants will be more powerful than the settlement encouraging effect of framing on plaintiffs, so the former is likely to predominate” (Rachlinski and Wistrich 2013).

When it comes to low-probability outcomes, however, we see a different pattern. People tend to be risk-seeking (rather than risk-averse) toward low-probability gains and risk-averse (rather than risk-seeking) toward low-probability losses (Kahneman and Tversky 1979). This means that in the domain of cases with a low likelihood of plaintiff winning—what some might call frivolous cases—plaintiffs, facing gains, are likely to be risk-seeking and prefer trial, while defendants, facing losses, are likely to become risk-averse and prefer settlement (Guthrie 2000; see also Fobian and Christensen-Szalanski 1993). Experimental evidence is consistent with this prediction—in the face of low odds, plaintiffs are expected to prefer trial and defendants to prefer settlement (Guthrie 2000). Additional research is needed to determine whether and to what extent such expectations translate to litigation decisions.

It is also the case that choices made by either side can be influenced by whether the litigant perceives a particular option as a gain or a loss—a perception that can be affected by the reference point to which the option is compared. For example, a plaintiff might compare a settlement offer to the expected value of going to trial and see the settlement as a gain, or the salient comparison might be to the plaintiff’s position prior to the injury in which case the same offer might be viewed as a loss. A settlement

option that is seen as a loss will be less attractive than the same settlement option that is perceived as a gain (Korobkin and Guthrie 1994b). It seems likely that the postinjury, pretrial situation will be the most intuitive frame for litigants in most instances (see Zamir and Ritov 2010). Nonetheless, alternative frames may be salient in certain cases.

The desire to avert losses has a variety of consequences for decision-making. One with particular consequence for settlement decision-making is the tendency to overweight *sunk costs*. Under the standard economic model, costs that have already been incurred and cannot be recouped should not influence decisions about whether to incur or invest additional resources. However, such prior costs have a tendency to influence decision-making (see Arkes and Blumer 1985). As a result, investments (of time or money) made at earlier stages of litigation can influence decisions at later stages, even when they should be irrelevant (see, e.g., Kiser 2010).

## 6 NONMONETARY INFLUENCES

---

Expected value calculations are also complicated by the fact that litigants are motivated by a variety of monetary and nonmonetary incentives—many of which can be difficult to quantify and alter litigant judgments in other ways (see generally Robbennolt and Sternlight 2012). For example, litigants may desire to receive an explanation or even an apology. They may worry about reputations or seek a public judgment of accountability or vindication. They may seek to effect reform or behavioral change in order to prevent future harm. They care about “fairness” and about achieving a “just” resolution of their dispute. They may desire some other symbolic outcome. Ultimately, litigants, like other legal decision-makers, must attempt to make settlement decisions that balance a range of varied (and sometimes inconsistent) goals (see Robbennolt, Darley, and MacCoun 2003; Robbennolt, MacCoun, and Darley 2010).

Consider, for example, that litigants are motivated to minimize the amount of regret that they will experience following their decisions. Regret is “the painful feeling a person experiences upon determining she could have obtained a better outcome if she had decided or behaved differently” (Guthrie 1999, pp. 65–66; see also Gilovich and Medvic 1995). Because feelings of regret can be painful, people attempt to predict such feelings and act so as to avoid them. This tendency is predicted to promote settlement. When litigants reject a settlement offer or demand and proceed to trial, they will be able to compare the rejected settlement to the outcome at trial, raising the possibility that they will regret their decision. Conversely, litigants who agree to a settlement will never know what might have happened at trial. The absence of this comparison minimizes the regret that they expect to experience (Guthrie 1999). Consistent with these predictions, Guthrie (1999) found that settlement was predicted to be more likely in a system in which the outcome of trial would not be known as compared to a hypothetical system in which both outcomes would be known. In this way, then, litigants may take into account not only the monetary value of the offer on the table and the predicted value

of any trial outcome, but also “the feelings associated with the outcomes of foregone options” (Guthrie 1999, p. 69).

At the same time, this desire to minimize regret is also associated with a tendency to make decisions that are geared toward keeping options open—that is, making choices that avoid foreclosing other options—even when such decisions perpetuate uncertainty and when they are costly (see, e.g., Shin and Ariely 2004). But such a strategy can get in the way of settlement. Rejecting a particular settlement offer forecloses that particular offer, but keeps open both the possibility of trial and the possibility of settlement. In addition, such decision-making can reduce the degree to which people are satisfied with their decisions (see Bullens, van Harreveld, and Förster 2011; Gilbert and Ebert 2002).

Regret about a forgone opportunity can also affect how a current proposal is evaluated. *Inaction inertia* is the “tendency of a person to omit action when he or she has passed up a similar, more attractive opportunity to act” (Anderson 2003, p. 146). This tendency can induce a litigant who has passed up a proposed settlement to compare any subsequent proposal to that previous proposal even though that option may now be off the table. “For example, imagine a defendant who passes up an offer to settle a case for \$10,000, thinking that she can push the plaintiff to settle for a lower figure. But imagine that the plaintiff does not settle for less and withdraws the offer to settle for \$10,000. Down the road, the defendant is likely to find settling for \$12,000 particularly unattractive, even if the defendant thinks it is likely that a trial would result in liability of \$15,000” (Robbennolt and Sternlight 2012, p. 94).

Settlement decisions can also be influenced by litigant views of the *procedural justice* that they have experienced or expect to experience. In addition to wanting to experience a neutral and trustworthy process, litigants care about having a voice in the process and about being treated with dignity and respect (see Blader and Tyler 2003). While litigants and their attorneys likely realize that the other side will be focused on looking out for its own interests, they still want the other side to listen, to be respectful in their interactions, and to act in good faith (see Hollander-Blumoff and Tyler 2008). Indeed, when negotiators act in a procedurally just manner, their predictions about likely outcomes tend to be less far apart and they tend to find common ground more quickly, are less likely to fail to reach agreement, and are more satisfied with the resulting agreement (Hollander-Blumoff and Tyler 2008; Leung, Tong, and Ho 2004; see also Okimoto and Tyler 2007).

Similarly, litigants are influenced by the perceived procedural justice of other settlement mechanisms. For example, researchers have explored the decisions of both individual and business litigants in cases involving a wide range of stakes that were sent to court-mandated nonbinding arbitration. Parties could accept the arbitrator’s award or go to trial. Decisions about whether to accept the arbitrator’s award were more strongly associated with perceptions of the procedural justness of the arbitration process than with the size of the arbitrator’s award (Lind et al. 1993).

Litigants may also be motivated by the desire to have their “day in court.” To take one example, Rob MacCoun describes



a judicial settlement conference in which the attorneys, with no clients present, hammered out a settlement they were comfortable with, but the plaintiff's attorney complained that his client might not accept it because she "wants her day in court." The judge put on his robe, called her into an empty courtroom, and sat her on the witness chair. After she told her story, she assented to the settlement. (MacCoun 2005, p. 189)

Finally, while not formally part of an expected value calculus, litigation can be forestalled by an apology or prompted when an actor fails to apologize. Similarly, settlement negotiations can be influenced by whether an apology is proffered and the nature of that apology (see Korobkin and Guthrie 1994b; Robbennolt 2003, 2006, 2008). Apologies can change litigants' perceptions, emotions, attributions of blame, and desire for retribution. Consequently, apologies can make it less likely that claimants will turn to the legal system following an injury (Mazor et al. 2004, 2006) and can influence the course of settlement negotiations (Robbennolt 2003, 2006).

## 7 ADVERSARIAL POSTURE OF LITIGATION

---

In litigation, the parties are typically styled as adversaries engaged in bargaining about a potential settlement. In an economic model that approaches litigation bargaining as a distributional problem, the parties are engaged in a process of bargaining over the litigation surplus (see generally Cooter, Marks, and Mnookin 1982). The "negotiator's dilemma," then, is that acting cooperatively—by, for example, disclosing information—can simultaneously serve to help the parties find integrative solutions that expand the surplus, but also make them vulnerable to exploitation by the other side (Lax and Sebenius 1986). This adversary posture and the negotiator's dilemma have a range of implications for whether and how cases are settled.

As noted above, there are a variety of strategic bargaining tactics that can distort decision-making—in some cases leading to trial when the parties would have been better served by a settlement and in other cases leading to inadvisable settlements. For example, one or both sides may engage in misrepresentation or bluffing, both sides will attempt to persuade each other to make concessions (through argumentation, threats, or other attempts to influence), or one or both sides may adamantly refuse to settle (to discourage litigation or for a host of other reasons). Parties may use discovery strategically. To the extent that settlement negotiation is about dividing a surplus, the parties may reject particular offers that would otherwise be acceptable in an attempt to gain a larger share. The disclosure or withholding of information can also be the basis of strategic decision-making.

But, in addition to these strategic aspects of bargaining, the fact that litigants are adversaries introduces a variety of other distortions of judgment. For example, people tend to attribute bias to those with whom they disagree—simply because they disagree—which leads them to be less inclined to cooperate and more inclined to use



competitive bargaining strategies (Kennedy and Pronin 2008). Similarly, adversaries tend to assume that their interests in settlement are incompatible—that is, that a gain for one side must necessarily be a loss for the other. This *fixed pie bias* can lead parties to emphasize the distributive aspects of the negotiation over the integrative aspects and lead them to miss mutually acceptable opportunities for settlement (Bazerman and Neale 1983; Thompson and Hastie 1990; see also Thompson and Hrebec 1996). Relatedly, proposals that are offered by an adversary often seem less attractive than the same proposal would have if offered by another source (Ross and Ward 1995; Korobkin and Guthrie 1994b). This *reactive devaluation* can lead parties to decline settlement that would otherwise be acceptable. Both sides, therefore, might be well served by a mediator or other intermediary who can be attentive to value-creating openings and present proposals as neutral in origin (see Robbennolt and Sternlight 2012).

Finally, behavioral studies have shown that adversaries can be motivated by the prospect of retribution and concerned with eliminating injustice through punishment (see Miller and Vidmar 1981). For example, people will incur costs to punish others when they believe it is appropriate (see deQuervain et al. 2004; Fehr and Gächter 2002; Henrich et al. 2006). But, of course, it is difficult for opposing sides to agree on the appropriate degree of punishment. In assessing the nature and consequences of a transgression, there is a *magnitude gap* such that punishments that seem fair to the injured party seem excessive to the transgressor and vice versa (see Stillwell et al. 2008). These differences in perceptions can make it difficult to reach settlement.

## 8 INFORMATION GATHERING

---

One set of tasks involved in litigating and attempting to settle cases entails determining how much information to seek. Importantly, obtaining additional information may make it possible to make better calibrated predictions about trial outcomes or predictions about the other side's bargaining behavior—arguably making settlement more likely. In some cases, the additional information obtained through discovery leads a plaintiff to re-evaluate and drop a case or a defendant to agree to a settlement (see Farber and White 1991; Golann 2011).

But more information is not a panacea. Obtaining additional information is costly in terms of time and money (see Pace and Zakaras 2012; see also Shepherd 1999), and does not always improve decision-making. First, decision-makers are not always well calibrated in deciding what information to seek. The *confirmation bias* can cause decision-makers to seek out information that confirms an existing belief or hypothesis, while ignoring potentially disconfirming evidence (Nickerson 1998). Indeed, the confirmation bias can lead lawyers to seek out superfluous information (and to miss relevant information) (see Rachlinski and Wistrich 2013). Confirming (and failing to disconfirm) beliefs in the merits of one's case is likely to decrease the bargaining range and make settlement more difficult. Failure to gather the appropriate

information can make for poor predictions and can make it difficult to accurately value options. In addition, irrelevant information can be distracting (Wittermoth and Neale 2011).

Second, any information gained by the parties is subject to interpretation. *Biased assimilation* leads people to interpret new information in ways that allows them to confirm, rather than challenge, existing beliefs (Lord, Ross, and Lepper 1979). Thus, “shared information, if it is open to multiple interpretations, is likely to be interpreted egocentrically by the disputants, which can cause beliefs to diverge rather than converge” (Loewenstein and Moore 2004, p. 37). Similarly, additional information can open up new benchmarks for understanding fairness, which can also be interpreted in self-serving ways (Camerer and Loewenstein 1993). And *naive realism*—the view that one’s own take on the world is objective and shared—makes it difficult to see how others might view the information differently (see Pronin, Puccio, and Ross 2002).

Third, seeking out information can take on a life of its own. In particular, there is a tendency to delay decisions until information is available, even if the sought after information would not really affect the decision. For example, people tend to want to obtain information about whether they would be punished before engaging in a prohibited act—even if they would have been inclined to engage in the act had they known that they would be punished (Bastardi and Shafir 2000; see also Redelmeier, Shafir, and Aujila 2001). Not only can attempting to obtain such (ultimately irrelevant) information be costly, but it can also change the nature of the decision at issue. Thus, in the study just described, most people who knew from the start that punishment would ensue chose to act anyway. On the other hand, when it was unclear whether they would be punished, most people chose to wait to decide until they could find out. When they then found out that punishment would follow from the act, they chose *not* to act (Bastardi and Shafir 2000). In the litigation context, consider a study in which lawyers were assigned to advise a fictional plaintiff about a settlement offer. Those who were told initially that a government safety report implicated the defendant’s machine were more likely to advise settlement than were lawyers who elected to wait for the report and only then discovered that it implicated the defendant. Waiting for the (same) information apparently gave it additional value, making the plaintiff’s prospects at trial seem brighter and the proposed offer less appealing (Rachlinski and Wistrich 2013). Seeking out information imbued that information with seeming importance that altered the decision (see also Young et al. 2012).

## 9 SETTLEMENT OPTIONS

---

Another behavioral complication is that decision-makers do not simply evaluate the expected value of options in isolation. Instead, decision-makers evaluate options in the context of other options that are, were, or could be on the table.

For example, when evaluating the pros and cons of an option in the context of other options, loss aversion means that negative features of the option tend to be more salient than positive features. This *option devaluation* means that people tend to find a given option less appealing when it is appraised in the context of other options that it would have been if considered alone (Brenner, Rottenstreich, and Sood 1999; Simonson and Tversky 1992). In addition, a *contrast effect* can occur when an option is similar but inferior to another option under consideration. The contrast between the similar options can magnify the appeal of the superior option even in comparison to other nonsimilar options (Guthrie 2003; Huber, Payne, and Puto 1982; Kelman, Rottenstreich, and Tversky 1996). Finally, a *compromise effect* occurs when the addition of an extreme option to the set of options under consideration makes other options appear more moderate and, therefore, more attractive (Guthrie 2003; Kelman, Rottenstreich, and Tversky 1996; Simonson and Tversky 1992).

Note also that decision-makers evaluate options differently depending on whether they are attempting to *accept* or to *reject* one or more options. When looking for an option to accept, the positive features of the options are front and center. In contrast, when thinking about which options to reject, the focus is on the downsides of the options. This means that an option for which there are both good reasons to accept it *and* good reasons to reject it can be both more likely to be selected *and* more likely to be rejected depending on the decision-maker's orientation to the decision (Shafir 1993; see also Rachlinski, Guthrie, and Wistrich 2009).

## 10 LAWYERS AS AGENTS

---

“[T]he feature of litigation bargaining that most differentiates it from other types of negotiation is the presence of lawyers” (Korobkin and Guthrie 1997, p. 81). Accordingly, any behavioral analysis of litigation and settlement ought to take into account the similarities and differences in how litigants and their lawyers make decisions, any factors that lead their incentives to diverge, and the dynamic relationship between lawyer and client. As we will see, the presence of lawyers may either facilitate settlement or impede it (see Gilson and Mnookin 1994; Sternlight 1999)—either of these possibilities has the potential to well- or ill-serve the client depending on the circumstances.

The interests of attorneys and clients will often converge—both clients and attorneys have an interest in successfully resolving a given case. But attorneys also have interests in generating fees, their own professional advancement, establishing or maintaining a particular reputation, and productively allocating their time across clients and other activities. Thus, there is ample room for the interests of lawyers and clients to diverge.

One factor that has received much attention is the effects of different fee arrangements on litigation and settlement. Indeed, there is a sizable body of economic literature that theorizes about the potential effects of how fees are structured and who pays them (see review in Katz 2000; Miller 1987). Recent behavioral studies have also

examined the factors that influence the attractiveness of different fee agreements (see Zamir and Ritov 2010 [loss aversion]; Zamir and Ritov 2011 [fairness constraints]).

Whether legal fees are structured based on time spent (e.g., hourly billing), a fixed price, or a contingent basis surely has implications for the incentives faced by attorneys in how to structure their work and caseloads. For example, hourly billing might motivate attorneys to bring lower quality cases, spend too much time on a matter, and to delay settlement; flat-fee or other fixed-rate billing systems might motivate the minimizing of effort spent on a matter, and a contingent fee system is likely motivate behavior aimed at maximizing the attorney's return on investment and is likely to give primacy to the financial aspects of settlement to the neglect of nonmonetary concerns (see Kritzer 2002; see also Helland and Tabarrok 2003). At the same time, however, it is likely that focusing primarily on these billing concerns is too narrow. Such concerns are likely just one of many considerations that influence how attorneys handle cases. Attorney decision-making under any fee arrangement is likely influenced by the range of alternative uses attorneys might have for their time (opportunity costs), reputational concerns, and the incentives created by the ongoing representation of a client and the importance of satisfied clients and referrals in generating business (Kritzer 2002). "For every plausible claim that lawyers' fee arrangements have one-directional effect on settlement rates, there is an equally plausible claim that lawyers' long-term financial incentives might cut in the opposite direction" (Korobkin and Guthrie 1997, p. 123). Attorneys must balance these competing considerations (see Robbennolt, MacCoun, and Darley 2010), and researchers ought to develop a better understanding of how attorneys do so.

But in addition to the complications of fee structures, there are also aspects of the ways in which litigants and attorneys make decisions and of the different approaches and abilities they bring to the litigation and settlement table that can influence settlement. There are a variety of ways in which we might expect attorneys' decision-making to differ from that of at least some of their clients. For example, clients likely know more about their own goals and preferences, while lawyers likely know more about the substance of the law and legal processes. Lawyers tend to be more analytical than many clients. This means that lawyers may be more likely to apply a version of expected value analysis than are clients—a mode of analysis that may lead them to either encourage or discourage settlement depending on the circumstances. But it can also mean that lawyers can have difficulty fully understanding clients' preferences and goals in ways that can make it hard for them to "bring the client" along toward settlement (see Korobkin and Guthrie 1997; Robbennolt 2008; Sternlight 1999). In addition, it can be difficult for lawyers to simultaneously adopt the roles of objective advisor and zealous advocate (Moore, Tanlu, and Bazerman 2010). And lawyers are more likely to be repeat players in the process, increasing both the importance and effects of a concern for reputation (see Gilson and Mnookin 1994; Rubin and Sander 1988).

Nonetheless, the extent to which attorneys' decision-making differs from that of their clients' is not entirely clear (nor is it clear how much variation there is among different clients or types of clients). There is evidence that attorneys (and other experts)

are influenced by many of the phenomena discussed in this chapter. But there is also some evidence that attorneys or other agents may be differently influenced by some of them (see, e.g., Korobkin and Guthrie 1997). Moreover, to the extent that there are differences in the decision-making processes of attorneys and clients, additional research on how these differences get negotiated between attorney and client to result in decisions throughout the litigation and settlement process is sorely needed.

## 11 CONCLUSION

---

A behavioral analysis of litigation and settlement reveals that litigants and their attorneys are likely to have difficulty behaving and deciding in ways that are consistent with a straightforward model of utility maximization. Predictions about expected outcomes can be influenced by a variety of phenomena, including self-serving biases, overconfidence, the lure of available anchors, and whether the decision-maker is facing a decision regarding gains or losses. Litigants' expectations about their future emotional states, desire for justice in its many forms, and interest in nonmonetary outcomes can complicate decision-making. Psychological reactions to adversaries, their goals, and their proposals can get in the way of value-maximizing settlements. Gathering new information cannot always be counted upon to narrow gaps in perceptions and can sometimes distort judgment. Expanding the set of options under consideration is often beneficial, but can have unexpected consequences. And the dynamic interaction between lawyer and client has implications for whether cases are settled and on what terms.

Many of these behavioral influences are likely to impede settlement. Overconfidence, reactive devaluation, and other phenomena are likely to decrease the bargaining range and make it more difficult for the two sides to find common ground. At the same time, however, there are some influences that are likely to push toward settlement. Adding an extreme option might make an existing proposal look attractive as a compromise, and procedurally just treatment may pave the way toward agreement. Neither set of pressures, however, is uniformly desirable or undesirable. As noted above, there are certainly times at which parties fail to reach beneficial settlements, but also instances of ill-advised settlements. Thus, it will be important for future research to clearly articulate the conditions under which various psychological factors are likely to impede or promote both quality and poor settlements.

In the meantime, lawyers and their clients likely need to engage in a counseling process that acknowledges the psychological influences on both lawyer and client, capitalizes on differences in their respective knowledge bases and approaches to decision-making, and aims to resist any counterproductive influences. Techniques such as considering the opposite (Lord, Lepper, and Preston 1984), taking an outside perspective (Kahneman and Lovallo 1993), and considering the perspective of a

disagreeable adjudicator (Korobkin 2006) can all be helpful in this regard (see generally Robbennolt and Sternlight 2012).

Appreciating the range of behavioral influences on decision-making in the context of litigation and settlement allows for a more nuanced understanding of the decisions that litigants and their attorneys must make throughout the litigation and settlement process. Continued empirical research into the ways in which litigants and their attorneys make predictions, set goals, weigh options, and interact with each other will further enrich this understanding.

## REFERENCES

- Anderson, C. J. 2003. The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psychological Review* 129:139–67.
- Arkes, H. R., and C. Blumer. 1985. The psychology of sunk cost. *Organizational Behavior and Human Decision Processes* 35:124–40.
- Babcock, L., H. S. Farber, C. Fobian, and E. Shafir. 1995a. Forming beliefs about adjudicated outcomes: Perceptions of risk and reservation values. *International Review of Law and Economics* 15:298–303.
- Babcock, L., and G. Loewenstein. 1997. Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives* 11:109–26.
- Babcock, L., G. Loewenstein, S. Issacharoff, and C. Camerer. 1995b. Biased judgments of fairness in bargaining. *American Economic Review* 85:1337–42.
- Babcock, L., and G. Pogarsky. 1999. Damage caps and settlement: A behavioral approach. *Journal of Legal Studies* 28:341–70.
- Bailis, D. S., and R. J. MacCoun. 1996. Estimating liability risks with the media as your guide: A content analysis of media coverage of tort litigation. *Law and Human Behavior* 20:419–29.
- Bastardi, A., and E. Shafir. 2000. Nonconsequential reasoning and its consequences. *Current Directions in Psychological Science* 9:216–19.
- Bazerman, M., and M. Neale. 1983. Heuristics in negotiation: Limitations to dispute resolution effectiveness. In *Negotiating in Organizations*, M. H. Bazerman and R. Lewicki. London: Sage, 51–67.
- Bebchuk, L. 1984. Litigation and settlement under imperfect information. *Rand Journal of Economics* 15:404–15.
- Blader, S. L., and T. R. Tyler. 2003. A four-component model of procedural justice: Defining the meaning of a “fair” process. *Personality and Social Psychology Bulletin* 29:747–58.
- Boyd, C. L., and D. A. Hoffman. 2013. Litigating toward settlement. *Journal of Law, Economics, and Organization*, 29:898–929.
- Brenner, L. A., D. J. Koehler, and A. Tversky. 1996. On the evaluation of one-sided evidence. *Journal of Behavioral Decision Making* 9:59–70.
- Brenner, L., Y. Rottenstreich, and S. Sood. 1999. Comparison, grouping, and preference. *Psychological Science* 10:225–29.
- Bullens, L., F. van Harreveld, and J. Förster. 2011. Keeping one’s options open: The detrimental consequences of decision reversability. *Journal of Experimental Social Psychology* 47:800–805.



- Camerer, C. F., and G. Loewenstein. 1993. Information, fairness, and efficiency in bargaining. In *Psychological Perspectives on Justice: Theory and Applications*, B. A. Mellers and J. Baron. New York: Cambridge University Press, 155–79.
- Cooter, R. D., S. Marks, and R. Mnookin. 1982. Bargaining in the shadow of the law: A testable model of strategic behavior. *Journal of Legal Studies* 11:225–51.
- Cooter, R. D., and D. L. Rubinfeld. 1989. Economic analysis of legal disputes and their resolution. *Journal of Economic Literature* 27:1067–97.
- Cooter, R. D., and T. Ulen. 2012. *Law and Economics*. 6th ed. Boston: Prentice Hall.
- Depoorter, B. 2010. Law in the shadow of bargaining: The feedback effects of civil settlements. *Cornell Law Review* 95:957–87.
- de Quervain, D. J. F., U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck, and E. Fehr. 2004. The neural basis of altruistic punishment. *Science* 305:1254–58.
- Eisenberg, T. and C. Lanvers. 2009. What is the settlement rate and why should we care? *Journal of Empirical Legal Studies* 6:111–46.
- Fehr, E., and S. Gächter. 2002. Altruistic punishment in humans. *Nature* 415:137–40.
- Felstiner, W. L. F., R. L. Abel, and A. Sarat. 1980–81. The emergence and transformation of disputes: Naming, blaming, claiming. . . *Law and Society Review* 15:631–54.
- Fobian, C. S., and J. J. Christensen-Szalanski. 1993. Ambiguity and liability negotiations: The effects of negotiators' role and the sensitivity zone. *Organizational Behavior and Human Decision Processes* 54:277–98.
- Fox, C. R., and R. Birke. 2002. Forecasting trial outcomes: Lawyers assign higher probability to possibilities that are described in greater detail. *Law and Human Behavior* 26:159–73.
- Galanter, M. 1986. The day after the litigation explosion. *Maryland Law Review* 46:3–39.
- Galanter, M. 2004. The vanishing trial: An examination of trials and related matters in federal and state courts. *Journal of Empirical Legal Studies* 1:459–570.
- Galinsky, A. D., and T. Mussweiler. 2001. First offers as anchors: The role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology* 81:657–69.
- Gilbert, D. T., and J. E. J. Ebert. 2002. Decisions and revisions: The affective forecasting of changeable outcomes. *Journal of Personality and Social Psychology* 82:503–14.
- Gilovich, T., and V. H. Medvec. 1995. The experience of regret: What, when, and why. *Psychological Review* 102:379–95.
- Gilson, R. J., and R. H. Mnookin. 1994. Disputing through agents: Cooperation and conflict between lawyers in litigation. *Columbia Law Review* 94:509–66.
- Golann, D. 2011. Dropped medical malpractice claims: Their surprising frequency, apparent causes, and potential remedies. *Health Affairs* 30:1343–50.
- Goodman-Delahunty, J., P. A. Granhag, M. Hartwig, and E. F. Loftus. 2010. Insightful or wishful: Lawyers' ability to predict case outcomes. *Psychology, Public Policy, and Law* 16:133–57.
- Gross, S. R., and K. D. Syverud. 1991. Getting to no: A study of settlement negotiations and the selection of cases for trial. *Michigan Law Review* 90:319–93.
- Gross, S. R., and K. D. Syverud. 1996. Don't try: Civil jury verdicts in a system geared to settlements. *U.C.L.A. Law Review* 44:51–64.
- Guthrie, C. 1999. Better settle than sorry: The regret aversion theory of litigation behavior. *University of Illinois Law Review* 1999:43–90.
- Guthrie, C. 2000. Framing frivolous litigation: A psychological theory. *University of Chicago Law Review* 67:163–216.
- Guthrie, C. 2003. Panacea or pandora's box?: The costs of options in negotiation. *Iowa Law Review* 88:601–53.



- Guthrie, C., and D. Orr. 2006. Anchoring, information, expertise, and negotiation: New insights from meta-analysis. *Ohio State Journal of Dispute Resolution* 21:597–628.
- Guthrie, C., J. J. Rachlinski, and A. J. Wistrich. 2001. Inside the judicial mind. *Cornell Law Review* 86:777–830.
- Guthrie, C., J. J. Rachlinski, and A. J. Wistrich. 2009. The “hidden judiciary”: An empirical examination of executive branch justice. *Duke Law Journal* 58:1477–530.
- Hannaford-Agor, P., and N. Waters. 2013. Estimating the cost of civil litigation. *Caseload Highlights* 20:1–8.
- Helland, E., and A. Tabarrok. 2003. Contingency fees, settlement delay, and low-quality litigation: Empirical evidence from two datasets. *Journal of Law, Economics, and Organization* 19:517–42.
- Henrich, J., et al. 2006. Costly punishment across human societies. *Science* 312:1767–70.
- Hensler, D. R., et al. 1991. *Compensation for Accidental Injuries in the United States*. Santa Monica, CA: Rand.
- Hogarth, R. M. 1989. Ambiguity and competitive decision making: Some implications and tests. *Annals of Operations Research* 19:31–50.
- Hollander-Blumoff, R., and T. R. Tyler. 2008. Procedural justice in negotiation: Procedural fairness, outcome acceptance, and integrative potential. *Law and Social Inquiry* 33:473–500.
- Huber, J., J. W. Payne, and C. Puto. 1982. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Consumer Research* 9:90–98.
- Kahneman, D., and D. Lovallo. 1993. Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science* 39:17–31.
- Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47:263–92.
- Kahneman, D., and A. Tversky. 1984. Choices, values, and frames. *American Psychologist* 39:341–50.
- Katz, A. W. 2000. Indemnity of legal fees. In *The International Encyclopedia of Law and Economics*, vol. 5, B. Bouckaert and G. DeGeest. Cheltenham, UK: Edward Elgar Publishing, 63–94.
- Kelman, M., Y. Rottenstreich, and A. Tversky. 1996. Context-dependence in legal decision making. *Journal of Legal Studies* 25:287–318.
- Kennedy, K. A., and E. Pronin. 2008. When disagreement gets ugly: Perceptions of bias and the escalation of conflict. *Personality and Social Psychology Bulletin* 34:833–48.
- Kiser, R. 2010. *Beyond Right and Wrong: The Power of Effective Decision Making for Attorneys and Clients*. Heidelberg: Springer.
- Kiser, R., M. A. Asher, and B. B. McShane. 2008. Let’s not make a deal: An empirical study of decision-making in unsuccessful settlement negotiations. *Journal of Empirical Legal Studies* 5:551–91.
- Korobkin, R. 2002. Aspirations and settlement. *Cornell Law Review* 88:1–61.
- Korobkin, R. 2006. Psychological impediments to mediation success: Theory and practice. *Ohio State Journal on Dispute Resolution* 21:281–327.
- Korobkin, R., and C. Guthrie. 1994a. Opening offers and out-of-court settlement: A little moderation may not go a long way. *Ohio State Journal on Dispute Resolution* 10:1–22.
- Korobkin, R., and C. Guthrie. 1994b. Psychological barriers to litigation settlement: An experimental approach. *Michigan Law Review* 93:107–92.
- Korobkin, R., and C. Guthrie. 1997. Psychology, economics, and settlement: A new look at the role of the lawyer. *Texas Law Review* 76:77–141.

- Kritzer, H. 2002. Lawyer fees and lawyer behavior in litigation: What does the empirical literature really say? *Texas Law Review* 80:1943–83.
- Langbein, J. H. 2012. The disappearance of civil trial in the United States. *Yale Law Journal* 122:522–73.
- Lax, D. A., and J. K. Sebenius. 1986. *The Manager as Negotiator: Bargaining for Cooperation and Competitive Gain*. New York: Free Press.
- Leung, K., K. K. Tong, and S. S. Ho. 2004. Effects of interactional justice on egocentric bias in resource decisions. *Journal of Applied Psychology* 89:405–15.
- Lind, E. A., K. T. Kulik, M. Ambrose, and M. V. de Vera Park. 1993. Individual and corporate dispute resolution: Using procedural fairness as a decision heuristic. *Administrative Science Quarterly* 38:224–51.
- Loewenstein, G., S. Issacharoff, C. Camerer, and L. Babcock. 1993. Self-serving assessments of fairness and pretrial bargaining. *Journal of Legal Studies* 22:135–59.
- Loewenstein, G., and D. A. Moore. 2004. When ignorance is bliss: Information exchange and inefficiency in bargaining. *Journal of Legal Studies* 33:37–58.
- Loftus, E. F., and W. A. Wagenaar. 1988. Lawyers' predictions of success. *Jurimetrics* 28:437–53.
- Lord, C. G., M. R. Lepper, and E. Preston. 1984. Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology* 47:1231–43.
- Lord, C. G., L. Ross, and M. R. Lepper. 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37:2098–109.
- MacCoun, R. J. 2005. Voice, control, and belonging: The double-edged sword of procedural fairness. *Annual Review of Law and Social Science* 1:171–201.
- Mazor, K. M., G. W. Reed, R. A. Yood, M. A. Fischer, J. Baril, and J. H. Gurwitz. 2006. Disclosure of medical errors: What factors influence how patients respond? *General Internal Medicine* 21:704–10.
- Mazor, K. M., S. R. Simon, R. A. Yood, B. C. Martinson, M. J. Gunter, G. W. Reed, and J. H. Gurwitz. 2004. Health plan members' views about disclosure of medical errors. *Annals of Internal Medicine* 140:409–18.
- Merry, S. E., and S. S. Silbey. 1984. What do plaintiffs want? Reexamining the concept of dispute. *Justice Systems Journal* 9:151–78.
- Miller, D. T., and N. Vidmar. 1981. The social psychology of punishment reactions. In *The Justice Motive in Social Behavior: Adapting to Times of Scarcity and Change*, M. J. Lerner and S. C. Lerner. New York: Springer, 145–72.
- Miller, G. P. 1987. Some agency problems in settlement. *Journal of Legal Studies* 16:189–215.
- Miller, R. E., and A. Sarat. 1981. Grievances, claims, and disputes: Assessing the adversary culture. *Law and Society Review* 15:52–62.
- Mnookin, R. H., and L. Kornhauser. 1979. Bargaining in the shadow of the law: The case of divorce. *Yale Law Journal* 88:950–97.
- Moore, D. A., and P. J. Healy. 2008. The trouble with overconfidence. *Psychological Review* 115:502–17.
- Moore, D. A., L. Tanlu, and M. H. Bazerman. 2010. Conflict of interest and the intrusion of bias. *Judgment and Decision Making* 5:37–53.
- Nickerson, R. S. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2:175–220.
- Okimoto, T. G., and T. R. Tyler. 2007. Is compensation enough? Relational concerns in responding to unintended inequity. *Group Processes and Intergroup Relations* 10:399–420.

- Pace, N. M., and L. Zakaras. 2012. *Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery*. Santa Monica, CA: Rand Institute for Civil Justice.
- Pronin, E., C. Puccio, and L. Ross. 2002. Understanding misunderstanding: Social psychological perspectives. In *Heuristics and Biases: The Psychology of Intuitive Judgment*, T. Gilovich, D. Griffin, and D. Kahneman. New York: Cambridge University Press, 636–65.
- Rachlinski, J. J. 1996. Gains, losses and the psychology of litigation. *Southern California Law Review* 70:113–85.
- Rachlinski, J. J. 2011. The psychological foundations of law and economics. *University of Illinois Law Review* 2011:1675–96.
- Rachlinski, J. J., C. Guthrie, and A. J. Wistrich. 2009. Context effects in judicial decision making, CELS 2009 Fourth Annual Conference on Empirical Legal Studies Paper. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1443596](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1443596).
- Rachlinski, J. J., and A. Wistrich. 2013. How lawyers' intuitions prolong litigation. *Southern California Law Review*, 86:571–635.
- Redelmeier, D. A., E. Shafir, and P. S. Aujila. 2001. The beguiling pursuit of more information. *Medical Decision Making* 21:374–81.
- Robbennolt, J. K. 2003. Apologies and legal settlement: An empirical examination. *Michigan Law Review* 102:460–516.
- Robbennolt, J. K. 2006. Apologies and settlement levers. *Journal of Empirical Legal Studies* 3:333–73.
- Robbennolt, J. K. 2008. Attorneys, apologies, and settlement negotiation. *Harvard Negotiation Law Review* 13:349–97.
- Robbennolt, J. K., J. M. Darley, and R. J. MacCoun. 2003. Symbolism and incommensurability in civil sanctioning: Decision makers as goal managers. *Brooklyn Law Review* 68:1121–58.
- Robbennolt, J. K., R. J. MacCoun, and J. M. Darley. 2010. Multiple constraint satisfaction in judging. In *The Psychology of Judicial Decision Making*, D. Klein and G. Mitchell. New York: Oxford University Press, 27–39.
- Robbennolt, J. K., and J. R. Sternlight. 2012. *Psychology for Lawyers: Understanding the Human Factors in Negotiation, Litigation, and Decision Making*. Chicago, IL: ABA Publishing.
- Robbennolt, J. K., and C. A. Studebaker. 2003. News media reporting on civil litigation and its influence on civil justice decision making. *Law and Human Behavior* 27:5–27.
- Ross, L., and A. Ward. 1995. Psychological barriers to dispute resolution. *Advances in Experimental Social Psychology* 27:255–304.
- Rottenstreich, Y., and A. Tversky. 1997. Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review* 104:406–15.
- Rubin, J. Z., and F. E. A. Sander. 1988. When should we use agents? Direct vs. representative negotiation. *Negotiation Journal* 4:395–401.
- Saks, M. J. 1992. Do we really know anything about the behavior of the tort litigation system—and why not? *University of Pennsylvania Law Review* 140:1147–292.
- Schwab, S. J., and M. Heise. 2011. Splitting logs: An empirical perspective on employment discrimination settlements. *Cornell Law Review* 96:931–56.
- Shafir, E. 1993. Choosing versus rejecting: Why some options are both better and worse than others. *Memory and Cognition* 21:546–56.
- Shepherd, G. B. 1999. An empirical study of the economics of pretrial discovery. *International Review of Law and Economics* 19:245–63.
- Shin, J., and D. Ariely. 2004. Keeping doors open: The effect of unavailability on incentives to keep options viable. *Management Science* 50:575–86.

- Simonson, I., and A. Tversky. 1992. Choice in context: Tradeoff contrast and extremeness aversion. *Marketing Research* 29:281–95.
- Songer, D. R. 1988. Tort reform in South Carolina: The effect of empirical research on elite perceptions concerning jury verdicts. *South Carolina Law Review* 39:585–605.
- Spier, K. E. 2007. Litigation. In *Handbook of Law and Economics*, vol. 1, A. M. Polinsky and S. Shavell. Amsterdam: Elsevier, 259–342.
- Sternlight, J. R. 1999. Lawyers' representation of clients in mediation: Using economics and psychology to structure advocacy in a nonadversarial setting. *Ohio State Journal of Dispute Resolution* 14:269–366.
- Stillwell, A. M., R. F. Baumeister, and R. E. del Priore. 2008. We're all victims here: Toward a psychology of revenge. *Basic and Applied Social Psychology* 30:253–63.
- Thompson, L. 1995. "They saw a negotiation": Partisanship and involvement. *Journal of Personality and Social Psychology* 68:839–53.
- Thompson, L., and R. Hastie. 1990. Social perception in negotiation. *Organizational Behavior and Human Decision Processes* 47:98–123.
- Thompson, L., and D. Hrebec. 1996. Lose-lose agreements in interdependent decision making. *Psychological Bulletin* 120:396–409.
- Thompson, L., and G. Loewenstein. 1992. Egocentric interpretations of fairness and interpersonal conduct. *Organizational Behavior and Human Decision Processes* 51:176–97.
- Trubek, D. M., et al. 1983. The costs of ordinary litigation. *U.C.L.A. Law Review* 31:72–127.
- Tversky, A., and D. Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5:207–33.
- Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211:453–58.
- Tversky, A., and D. Kahneman. 1982. Judgment under uncertainty: Heuristics and biases. In *Judgment under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, and A. Tversky. New York: Cambridge University Press, 3–20.
- Tversky, A., and D. J. Koehler. 1994. Support theory: A nonextensional representation of subjective probability. *Psychological Review* 101:547–67.
- van Koppen, P. J. 1990. Risk taking in civil law negotiations. *Law and Human Behavior* 14:151–65.
- Wittermoth, S., and M. A. Neale. 2011. Too much information: The perils of non-diagnostic information in negotiations. *Journal of Applied Psychology* 96:192–201.
- Young, M. J., C. W. Bauman, N. Chen, and A. Bastardi. 2012. The pursuit of missing information in negotiation. *Organizational Behavior and Human Decision Processes* 117:88–95.
- Zamir, E., and I. Ritov. 2010. Revisiting the debate over attorneys' contingent fees: A behavioral analysis. *Journal of Legal Studies* 39:245–88.
- Zamir, E., and I. Ritov. 2011. Notions of fairness and contingent fees. *Law and Contemporary Problems* 74:1–32.

## CHAPTER 25

---

# BEHAVIORAL ECONOMICS AND PLEA BARGAINING

---

RUSSELL COVEY

### 1 INTRODUCTION

---

IN recent years, criminal law scholars have begun to apply the insights of behavioral economics to the study of plea bargaining with important results. Combined with critical work demonstrating a variety of systemic defects that undermine the market-efficiency thesis, this research suggests that conventional law and economics descriptions of plea bargaining as the product of rational bilateral exchange of entitlements driven by punishment-maximization/minimization strategies simply do not provide an adequate account of the plea-bargaining system (Bibas 2004, p. 2467). Although this research provides a powerful basis to reject the assumption that the outcomes of plea bargaining are cast fairly and accurately in “the shadow of the law,”<sup>1</sup> it also gives rise to a puzzle. Most of the cognitive quirks and biases identified by researchers, such as loss aversion, overconfidence, overdiscounting, and self-serving bias suggest that defendants should be consistently disinclined to plead guilty and that prosecutors and defendants should consistently disagree about what constitutes a fair price for a guilty plea. Accordingly, predictions about plea bargaining based only on behavioral economics might lead one to expect that plea bargaining would be a rare occurrence. Of course it is not. Plea bargaining is far and away the predominant procedural mechanism for the resolution of criminal charges. Criminal trials, not plea bargains, are the oddity. Thus, in the United States, 97 percent of federal convictions and 94 percent of state convictions are the result of guilty pleas (Department of Justice 2009). And while plea bargaining is

<sup>1</sup> The insight that bargaining occurs in the “shadow of law” was first developed in the context of divorce (Mnookin and Kornhauser 1979), and has been pursued in many different legal domains.

far less common in other nations, it appears to be steadily on the increase. Scholars have estimated that in Germany, for instance, approximately one in every four trials is settled (Langer 2004, p. 42 n. 196). In Italy, in the years from 1990 to 1998 serious crimes brought in the *tribunale* were resolved by use of the *patteggiamento*, a negotiated sentencing procedure, in 34 to 42 percent of cases (Langer 2004, pp. 52–53). In Poland, voluntary settlement of criminal charges rose from less than 8 percent of all adjudicated cases in 2002 to more than 40 percent of cases by 2005 (Luna and Wade 2010).

Much of the pioneering work in this area accepts plea bargaining as a given and examines the impact that cognitive bias has on the bargaining decisions made by prosecutors and defendants. That work concludes that cognitive bias sometimes impedes defendants from accepting utility-enhancing plea offers, and less frequently induces defendants to accept utility-diminishing plea bargains. This chapter considers the relationship between cognitive bias and plea bargaining, exploring why parties sometimes fail to agree to utility-enhancing bargains, and perhaps more importantly, why plea bargaining is so prevalent notwithstanding the existence of plea-discouraging cognitive bias.<sup>2</sup> This chapter shows not only how the criminal justice system is functionally designed to induce defendants to plead guilty, but also how many apparently arbitrary and oppressive features of current criminal practice can be explained largely as devices that function, in whole or in part, to neutralize the plea-discouraging effects of cognitive bias. Incorporation of the insights of behavioral economics into plea-bargaining theory thus provides a more nuanced explanation of the shape of many features of the criminal justice system. Behavioral economics also casts new light on the factors that drive plea-bargaining outcomes, including the magnitude of sentencing differentials, the pervasiveness of pretrial detention, and the prosaic procedural brutality that is a universal feature of virtually every individual encounter with the criminal justice system.

The chapter begins by examining several findings of behavioral economics and evaluates the likely impact of those findings on plea bargaining. These findings suggest that without more, cognitive bias, if left unchecked, should discourage resolution of criminal cases through negotiated settlement contrary to the predictions of simple economic analysis. The chapter then assesses a variety of features of the criminal justice system that work to check these tendencies.

<sup>2</sup> The basic framework in this chapter is drawn from previous work by the author (Covey 2007).



## 2 COGNITIVE BIAS AND PLEA-BARGAINING

---

### 2.1 Bounded Rationality

Bounded rationality refers to the relatively obvious proposition that human cognitive abilities are not unlimited. Because rationality is bounded, people cannot assess with perfect accuracy the costs and benefits of various courses of action. In general, people do not even attempt to maximize utility when making decisions. Instead, they “satisfice,” that is, once they have identified an option that appears “good enough,” they discontinue their search and select that option (Simon 1992; Gigerenzer 2012, § 2.1).

Bounded rationality presents a problem for the economic model of plea bargaining, which requires prosecutors and defendants alike to make rational calculations regarding the value of guilty pleas based on a variety of difficult to predict inputs. According to conventional economic theory, the value of a plea bargain is determined by reference to opportunity cost. The opportunity cost of a guilty plea is forfeiture of the right to a trial. Therefore, the value of a guilty plea depends on the expected value of trial (Reinganum 1988, p. 714; Boari and Fiorentini 2001, p. 222). The expected value of a guilty plea can be calculated pursuant to the formula  $V_p = (P \times Et) - R$ , where  $V_p$  represents the value of the guilty plea,  $P$  is the probability of conviction,  $Et$  is the expected sentence upon conviction at trial, and  $R$  is the resource cost of trial. While both plea bargains and trials entail resource costs, trials usually consume substantially more resources—both in terms of pretrial preparation and court-time, and often, postconviction review—than guilty pleas. Accordingly,  $R$  here represents the marginal resource costs of trial above those expended for guilty pleas. Applying this pricing model, rational defendants should prefer plea bargains whenever the expected punishment resulting from a guilty plea ( $V_p$ ) is less than the expected trial sentence multiplied by the probability of conviction, minus the added resource costs (if any) of trial.

Although relatively simple to calculate in controlled situations, making actual decisions based on this formula in the context of plea bargaining involves numerous uncertainties. First, the parties must generate an estimate of the probability that the defendant will be convicted at trial. Then they must predict what punishment will be imposed if the defendant is convicted at trial, estimate all of the costs involved in litigation, including attorneys’ fees, time lost waiting in court, and the psychological stress of nonresolution. They must then calculate the estimated trial sentence by multiplying the expected trial sentence by the probability of conviction and discount that figure by the estimated process costs in order to arrive at an understanding of what a guilty plea is actually “worth.” They then must predict what punishment will be imposed if they enter a guilty plea. Finally, they must compare those two values in order to decide which course of action to select.



Given the complexity of these tasks and the information deficits typical during the plea-bargaining process, boundedly rational defendants will satisfice rather than maximize. Probabilities of conviction depend on a wide assortment of factors that are hard to predict. In legal systems such as the United States, where discovery in criminal cases is restricted, defendants may not know what evidence the state has gathered, and even if they do gain knowledge of it, they still must anticipate how persuasive that evidence will be to a jury. If there are witnesses, they won't know whether the witnesses will show up on the day of trial, or what they will say if they do, or whether the testimony given will be credible. They cannot know in advance who will be on the jury, and thus they cannot estimate how sympathetic the jury might be to their case. At the same time, where the sentencing system identifies a range of punishment rather than a specific sentence, they may have only a general notion of the sentence they will actually receive upon conviction. Sentencing indeterminacy is not necessarily diminished in plea bargaining either because in many cases defendants are not offered a specific sentence in exchange for a guilty plea but instead a charge concession or a sentence recommendation, and therefore they often may have only a fuzzy notion of the likely consequences of entering a guilty plea, making any comparative assessment of plea bargaining as opposed to trial even more difficult.

In short, the plethora of variables, the absence of reliable information, and the difficulty of making accurate predictions create substantial obstacles to careful, precise, and accurate decision-making in plea bargaining. It is not immediately obvious, however, how this uncertainty should impact plea bargaining. Uncertainty by itself could make plea bargaining either more or less likely depending on the parties' respective risk tolerance. Persons who are risk-averse should view uncertainty as a "disutility" and thus be more likely to pursue a risk-minimization strategy, while persons who prefer risk should do the opposite. However, consideration of several additional well-documented cognitive biases working in combination with the high level of uncertainty inherent in criminal litigation creates a situation where boundedly rational defendants should disproportionately disfavor guilty pleas in favor of trials.

## **2.2 Overconfidence and Self-Serving Biases**

Cognitive research demonstrates that when it comes to evaluating risk, people are consistently overconfident (Babcock and Loewenstein 2000; p. 356; chapter 13 by Williams in this volume). Most people believe that they are above-average drivers, healthier, smarter, and more ethical than other people, and that good things will happen to them more often and bad things less often than statistics predict. In addition, people typically evaluate information selectively, giving more attention and credence to information that is consistent with their preexisting beliefs. This tendency is referred to as "confirmation," "egocentric," "optimism," or "self-serving" bias, and has been well documented by researchers (Teichman 2011; Weinstein 1980; Glockner and Engel 2012 [discussing how predefined roles, such as that of prosecutor or defendant, skew actors

behavior by creating biases in favor of one's position]). Research indicates that this bias is not easily countered. People persist in their inaccurate beliefs even where they are expressly informed of the probabilities and educated about self-serving bias, and even when faced with incentives to change their views (see Babcock and Loewenstein 2000, p. 361 [describing results of experiment showing that informing participants about self-serving bias did not affect tendency to interpret ambiguous data in a self-serving way]).

The expected effect of these biases on plea bargaining seems predictable. Overconfident criminal defendants should consistently overestimate the probability that they will prevail at trial. Prosecutors and defendants might often look at the same evidence but have divergent views as to its persuasiveness to a judge or jury. As a result, the parties should disagree about many or all of the inputs to the plea-pricing formula, including the probability of conviction, the trial sentence, and the plea sentence. Because self-serving bias makes it more difficult for the parties to reach a compromise that both sides perceive to be mutually beneficial, the predicted result is overlitigation of disputes, a prediction that appears to hold in the civil context (Rachlinski 2003, p. 1192).

### 2.3 Loss Aversion and Risk Aversion

Another well-documented cognitive phenomenon that is in tension with plea-bargaining is loss aversion. Loss aversion describes the tendency of actors to weigh losses more heavily than gains (Kahneman, Knetsch, and Thaler 1991, pp. 197–98; chapter 11 by Zamir in this volume). Prospect theory further suggests that different attitudes toward risk in the domains of gains and losses bias actors toward particular types of gain-protecting and loss-shedding conduct. Specifically, prospect theory predicts that people will protect gains by shunning risk while simultaneously preferring risk where it promises a chance to avoid taking losses (Kahneman and Tversky 1979). Because a plea bargain is a trade of a certain loss (a conviction and punishment) for a chance of no loss (an acquittal) accompanied by a chance of a greater loss (a trial conviction and enhanced sentence), the behavioral economics research suggests that all else being equal, we should expect people to gamble on total exoneration at trial rather than accept a certain, though likely smaller, punishment by pleading guilty (Birke 1999). It also supports a prediction that while defendants should seek out risk in order to avoid taking losses, prosecutors' plea-bargaining conduct will be comparatively conservative, since their incentives are reversed.

The documented fact that people tend to be less averse to loss-avoiding than gain-maximizing risk is paralleled by the more general observation that people's risk preferences differ. Rational choice theory assumes either indifference to risk or risk-aversion, but some people—gamblers and mountain climbers, for instance—affirmatively seek out risk (at least within delimited spheres of their lives)—and preferences toward risk can vary for individuals depending on context. Because a trial represents

risk, the tendency of many people to avoid risk might lead them to avoid trials in favor of plea bargains. However, because risk aversion decreases where risky actions carry the potential of averting losses, the general tendency toward risk aversion might be outweighed by risk-seeking in the domain of losses in the context of resolving serious criminal charges. Moreover, since criminals as a class would seem to be more risk-seeking than the general population, criminal defendants' might plausibly have particularly strong preferences for trials over plea bargains.

Loss aversion is likely related to another deep-seated cognitive characteristic: the endowment effect (see generally chapter 12 by Korobokin in this volume; chapter 11 by Zamir in this volume; Jolls and Sunstein 2006, p. 205). Conventional economic theory assumes that individuals' preferences are fixed and that the assignment of entitlements does not affect subjective valuations. Research has indicated that this assumption is unrealistic. The endowment effect describes the phenomena documented in several studies by which individuals consistently place a higher valuation on goods they possess than on those they do not.

Criminal defendants as a class likely possess fewer endowments than most, but the one asset the system provides is the right to a trial. That right is intuitively and widely understood. A defendant charged with a serious criminal offense should be expected to place a high valuation on the right, particularly once the prospect of a trial becomes tangible. This increase in valuation should necessarily increase the defendant's subjective valuation of the worth of his right to trial, increasing his general reluctance to relinquish it.

## 2.4 Overdiscounting

Related but distinct from risk preference is time preference, or discounting. Some discounting is plainly rational. *Ceteris paribus*, it is always marginally better to consume a good now (or defer a bad until later) than to defer gratification until tomorrow (or suffer the bad consequence now), for the simple reason that tomorrow may never come. However, while some discounting is rational, significant discounting may not be. People who overdiscount by disproportionately valuing present utility over future utility make a cognitive error. Some are aware of this tendency, and others are not. A person who fails to perceive the time preference error is said to be "overdiscounting." A person who understands the time preference error but cannot delay gratification, notwithstanding a desire to do so, is said to manifest "bounded willpower." In both cases, actors' choices do not maximize expected utility in the manner that rational choice theory predicts.

Just as there is evidence that predicts that criminals as a class are likely to be less risk-averse than others, research indicates that they are also more likely to overdiscount and to display bounded willpower (Jolls, Sunstein, and Thaler 1998, p. 1479). As a result, most persons charged with crimes should be expected to place a greater-than-average value on the short-term consequences of their actions and a smaller-than-average

value on risks that will manifest only in the long term. The emphasis on the near horizon thus would seem to weigh in favor of trials that carry the promise of near-term freedom, rather than plea bargains that guarantee near-term punishment.

## 2.5 Fairness Bias

“Fairness” bias might also be expected to lead individuals to plea bargain in ways not predicted by rational choice theory. Research into perceptions of fairness demonstrates a widespread willingness on the part of research subjects to forgo their rational self-interest in response to perceived unfair treatment. That people regularly allow considerations of fairness to override utility maximization has been demonstrated in experiments involving the “ultimatum game,” in which player A is given a sum of money (say, \$20), and must then divide that sum with player B. Player B has no control over the size of the shares, but can reject any proposed distribution, in which case neither party gets to keep the money. Straight rational choice theory suggests that Player A will propose a division giving him almost all of the \$20 and Player B very little (say, \$19 to \$1, or even \$19.99 to \$0.01), and that Player B will accept that proposal, since, *ex ante*, that strategy maximizes both players expected utility. After all, Player A maximizes expected utility by keeping as much of the pot as possible, and even when offered a small share, Player B is still better off taking it than rejecting it. Studies of the ultimatum game in practice, however, indicate that low offers are consistently rejected despite the fact that doing so fails to maximize expected utility (Roth 1995, pp. 253, 258; Korobkin and Ulen 2000, pp. 1135–38). Although such conduct is not consistent with the assumption of simple expected utility maximization, it is explainable by reference to considerations of fairness. That is, in some situations, people will willingly sacrifice their own short-term utility to avoid being treated “unfairly,” or to punish the other party for acting unfairly. Non-utility-maximizing norms, such as fairness, thus constrain bargaining parties and the strategies that are rationally available to them.

Given widespread perceptions among defendants that the criminal justice system does not treat them fairly, the fairness bias might be expected to encourage some defendants—especially, but not only, factually innocent ones—to reject a utility-enhancing plea offer when the outcome strikes them as unjust or if doing so appears to punish prosecutors who make unfair offers (Camerer and Thaler 1995). Indeed, some commentators have found evidence that defendants who perceive themselves as innocent will often reject utility-enhancing plea offers, to their detriment, because of their innocence. While some such defendants refusing to plead guilty win acquittals as a result of their refusal to plead guilty, others are convicted at trial and receive harsher trial sentences as a result (Gazal-Ayal and Tor 2011; Covey 2013).

Like self-serving bias, loss aversion, overdiscounting, and bounded willpower, fairness bias distorts outcomes predicted by expected utility maximization theory. Behavioral economics thus depicts a bargaining context dominated by converging vectors almost all of which point away from negotiated settlements and in the direction of

adversarial dispute resolution. Although each of these phenomena by itself may only marginally impact plea-bargaining decisions, taken together, they create powerful incentives for defendants' to reject guilty pleas.

Granted, there may be cognitive biases that operate in the contrary direction. Ambiguity-aversion might push some criminal defendants to prefer even a sure loss in the form of a plea bargain to an unknown loss at trial. Fairness bias might compel some defendants to prefer a plea to a trial if they expected unfair treatment by the judge or jury.

In addition, prosecutors' own preferences and cognitive biases may further work to counteract this collection of behavioral characteristics in encouraging guilty pleas. For example, prospect theory suggests that, just as criminal defendants might be especially reluctant to accept a plea deal that requires them to forfeit the chance for acquittal and take a certain loss, prosecutors should be more eager to protect the chance of a certain conviction—which represents a gain from the status quo—even if the cost is a more lenient sentence. Accordingly, we should expect to see fairly large sentencing discounts accompanying most plea deals, which, as will be discussed later in the chapter, we in fact do. Indeed, as the next section discusses, large sentence discounts are just one of numerous institutional features of the criminal justice process that are especially well designed to address whatever cognitive resistance to plea bargaining criminal defendants experience.

### **3. FEATURES OF THE CRIMINAL JUSTICE SYSTEM THAT OVERCOME COGNITIVE RESISTANCE TO PLEA BARGAINING**

---

Given the many well-documented cognitive tendencies that would seem to push criminal defendants to reject guilty pleas, one might reasonably expect that exercising the right to a jury trial would ostensibly be the preferred choice of criminal defendants. To summarize, a plea bargain asks overconfident, risk-preferring individuals to agree to suffer certain adverse consequences while abandoning trial rights with which they are endowed and which offer the only possible path to a loss-free outcome. In addition, the increased risk assumed by going to trial involves risk of added punishment that would not be consumed until some future date—perhaps one far off into the future. What is more, because most criminal defendants, at least in the United States, do not pay their own legal costs but instead receive state-compensated lawyers, one of the most significant economic incentives to compromise litigation—monetary legal costs—frequently is omitted from the defendant's calculus. It would seem, therefore, that plea bargains should be rare events. Obviously, they are not.

How, then, does the criminal justice system in the United States, and increasingly in other legal systems as well, manage to induce so many defendants to relinquish

their trial rights and accept guilty pleas in light of what would seem to be strong cognitive biases against admitting guilt? The answer, in short, is *by every means possible*. All things being equal, cognitive biases may well impel defendants to opt for trial, but things are far from equal. Most criminal justice systems contain numerous levers to induce defendants to abandon their right to trial and to accept guilty pleas, and their evolution has tended, with few exceptions, to expand and strengthen these levers. The evolutionary tendency of criminal justice systems toward increasing plea bargaining, especially in the United States, has long been apparent and has been thoroughly documented by plea-bargaining scholars such as Albert Alschuler (1975, 1981) and George Fisher (1993). What has not been fully appreciated, however, is the way the various features of the criminal justice system function as counterweights to the cognitive biases that would otherwise impede plea bargaining.

The next part of the chapter, which focuses on practices in the United States, describes several of these counterweights, starting with the most obvious and, undoubtedly, the most powerful: the simple fact that in the vast majority of cases, defendants who plead guilty get outcomes that are not only better than can be expected at trial, but better by orders of magnitude.

### 3.1 Overcoming Loss Aversion and Bounded Rationality: High Plea Discounts and Punitive Trial Penalties

There are several explanations for the puzzling persistence of extraordinarily high plea rates in light of the natural tendency of decision-makers to take risks to avoid certain losses. One explanation for why criminal defendants plead guilty in such great numbers is that criminal defendants are getting such “good deals”—that is, plea bargains in which the “utility value of the plea” is so much higher than that of trial—that even loss-averse defendants cannot say no (Zeisel 1980).

There is no definitive evidence establishing just how large the typical plea discount/trial penalty in the United States really is. But what evidence is available suggests that it is very large indeed. State sentencing statistics for 2002 show the median sentence for all felony cases adjudicated by guilty pleas to be 36 months, and the median sentence following jury trial conviction at 120 months, a 333 percent trial penalty. The mean statistics are similar: 52 months for guilty pleas, and 140 months for jury trials, which equates to a 292 percent trial penalty (Bureau of Justice Statistics 2002, at table 4.5: Average felony sentence lengths in State courts, by the type of conviction, type of sentence imposed, and offense).

Of course, raw statistics comparing trial sentences to plea sentences might miss important factors that account for some of the disparity. Perhaps the cases that go to trial are consistently more serious than the cases that plead out. Perhaps the opposite is true; the raw statistics provide no answer to the question. A leading study of plea discount rates used regression analysis to isolate the impact of method of conviction

on severity of sentence. Drawing on a limited data pool, the study found consistent evidence of a trial penalty ranging in size from roughly 13 percent to 461 percent, depending on state and offense of conviction (King et al. 2005, p. 992). Others have cited average state trial penalties of 300 percent, rising in some states to as high as 500 percent (McCoy 2003, p. 23). Abundant anecdotal evidence indicates that even higher trial penalties are not uncommon, particularly in sentences disposing of first-time offenders eligible for probation or other nonincarcerative dispositions. Substantial trial penalties can be manufactured merely by credible application of the statutory sentencing range for an offense, which is often expansive. Where the judge participates in plea bargaining, as is the practice in some jurisdictions, the credibility of the implied threat to impose a much stiffer sentence after conviction than offered before trial is undoubtedly substantial.

But large as they are, these figures generally fail to capture the actual trial penalties at work because they only compare the disparity of outcomes applying to the *same* offense of conviction. That, however, is the wrong comparison. In most cases, by pleading guilty a defendant not only bargains for the opportunity to receive a more lenient sentence for the offense of conviction, he also receives the opportunity to plead guilty to a less serious charge carrying a less onerous penalty. To calculate the actual plea discount in any particular case, therefore, one must compare the sentence imposed on the lesser charge to which the defendant pleaded guilty with the sentence that would have been imposed after conviction on potentially higher charges at trial.

An accurate estimate of the operative trial penalty, therefore, depends not only on raw sentence differentials but also on the amount and type of charge dismissal and movement that accompanies typical plea bargains. Although it is quite difficult to estimate the typical additional discount resulting from such charge reductions, that additional discount need not itself be large to have large effects. A 50 percent discount resulting from a charge reduction, for example, would magnify the average 300 percent trial penalty to 600 percent (and the “high-end” penalties to 1000 percent or more). Fifty percent does not seem to be an unrealistic discount resulting from charge bargaining. As a point of comparison, according to 2005 data compiled by the Virginia Sentencing Commission the difference in Virginia in average sentence for first-degree versus second-degree murder is about 50 percent.<sup>3</sup> Although we lack comprehensive data on charge movement patterns, it is easy to see what is possible. In a typical case described by one New York defense attorney, the defendant faced burglary charges carrying a statutory maximum term of 30 years. By negotiating a plea to unlawful entry (a lesser included offense, essentially, burglary minus intent to commit a felony, that carried a maximum sentence of six months), the lawyer won (at least on paper) a sixty-fold sentence discount (Kunen 1983, p. 95). Similarly, a typical plea bargain in a homicide case might involve an agreement to plead guilty to a second-degree murder charge

<sup>3</sup> Virginia Criminal Sentencing Commission, 2005 Annual Report, pp. 47–48 and figs. 40–42, available at <http://www.vcsc.virginia.gov/2005FULLAnnualReport.pdf> (last visited February 1, 2013).



in exchange for dismissal of a first-degree murder charge. The sentencing differential between first- and second-degree murder can easily approach or exceed 100 percent, as was true according to reported statistics in Virginia.<sup>4</sup> A murder charge reduced to manslaughter represents an even greater reduction, one that in federal court averages approximately 650 percent.<sup>5</sup> What data we do have, moreover, indicates that plea bargains entailing these sorts of charge reductions are fairly routine (Wright and Engen 2007).

Prosecutors possess a wide array of tools enabling them to reduce charges in ways likely to result in substantially enhanced plea discounts. Not only may prosecutors simply permit defendants to plead guilty to offenses carrying significantly reduced sentences, they may select charges tactically to expand the sentence differential when doing so provides desired bargaining leverage. Prosecutors may and do, for example, offer defendants with criminal histories an opportunity to avoid a career criminal “three strikes” sentence in exchange for a guilty plea. Empirical data demonstrates that prosecutors evade three-strikes laws by disproportionately charging lesser offenses against offenders who otherwise would be subject to three-strikes sentences. Prosecutors can also substitute a charge not subject to a mandatory minimum for one that is subject to such minimums, they can “stack” charges carrying mandatory minimums in order to threaten or impose dramatic increases in mandatory sentences after a trial conviction, or make available safety-valve provisions that waive statutory minimums to defendants who plead guilty. In states like California, the penal code gives prosecutors flexibility to charge some crimes either as felonies or misdemeanors (*Ewing v. California*, 538 U.S. 11 (2003)). Where a defendant commits a so-called wobbler offense, the difference between going to trial and pleading guilty might mean the difference between a felony theft conviction carrying substantial prison time and a misdemeanor theft conviction and probation. These tools only bolster other mechanisms, such as cooperation provisions, that permit prosecutors directly to obtain discounted sentences for defendants who plead guilty.

Sufficient data thus exist to suggest that plea discounts of a magnitude likely to overcome significant cognitive biases are common, if not routine. Such large differentials go a long way toward explaining how prosecutors manage to induce more than 90 percent of loss-averse criminal defendants to relinquish their trial rights and plead guilty.

These large plea discounts also help counteract other cognitive biases as well. Boundedly rational defendants may have substantial difficulty estimating whether the probability of conviction in their case is 50 percent or 90 percent, and thus might well be quite insensitive to marginal differences in plea offers. However, where defendants are offered the opportunity to avoid a 600 percent trial penalty, even substantial

<sup>4</sup> See *supra*, Virginia Criminal Sentencing Commission, 2005 Annual Report.

<sup>5</sup> See United States Sentencing Commission, Statistical Information Packet, Fiscal Year 2006, table 7 (showing that national federal median sentence for manslaughter is 37 months, and for murder 240 months).

variances in estimated probabilities fail to undermine the rational inducement to plead guilty. Rationality may be bounded, but it is usually not inoperative. Large sentencing differentials dramatically reduce ambiguity by exaggerating the penal consequences of the choice to contest a criminal charge, and thus make it easier for even boundedly rational and loss-averse decision-makers to make a utility-enhancing decision to plead guilty.

In short, there is ample evidence to support the argument that criminal defendants routinely receive extremely “good deals.” Although a variety of cognitive biases undoubtedly lead defendants to prefer trial over guilty pleas, sentencing differentials that make guilty pleas look not only better, but better by orders of magnitude go a long way toward helping to overcome them. Even strongly loss-averse defendants start to doubt the wisdom of holding out for trial in the face of these enormous differentials.

Still, “good deals” do not provide the whole explanation for the high plea rate. After all, such deals are not available in all cases, and defendants’ resistance to pleading guilty is not only a product of loss aversion and bounded rationality, but of a whole range of cognitive traits that work in concert to inhibit plea bargaining. An explanation for the high plea rate thus requires consideration of several additional facets of the criminal justice system—especially as manifested in large urban jurisdictions—that further counter cognitive biases that might otherwise lead to insistence on trial.

### 3.2 Overcoming Self-Serving Bias through Minimization of Outcome Uncertainty

The rational actor’s choice to plead guilty turns not only on the assessment of the potential penalty upon conviction, but also on the likelihood of conviction at trial. Substantial uncertainty regarding the probability of conviction, like uncertainty regarding the consequences, may well encourage loss-averse defendants to gamble on trial. Moreover, behavioral economics predicts that as a result of self-serving bias, confirmation bias, and overconfidence, defendants will harbor unrealistically positive assessments both of their chances of acquittal and of the likelihood of a light sentence. This section reviews several features of modern criminal process that work to blunt those tendencies.

#### 3.2.1 *Minimizing Probability-of-Conviction Uncertainty*

In the United States, access to discovery in criminal cases is far more constrained than it is in civil cases. Indeed, the US Supreme Court has flatly stated that “the Constitution does not require the prosecutor to share all useful information with the defendant.”<sup>6</sup> As a result, parties in criminal cases have less information about the evidence that will come out at trial than parties in civil cases. This information deficit

<sup>6</sup> See *United States v. Ruiz*, 536 U.S. 622, 629 (2002); see also *Weatherford v. Bursey*, 429 U.S. 545, 559 (1977) (“There is no general constitutional right to discovery in a criminal case”).

impedes plea bargaining by making it more difficult to predict trial outcomes. Limited discovery in criminal trials is a long-standing feature of criminal law that reflects and responds to the fundamental asymmetries of criminal litigation. Whereas in civil litigation, both parties have broad and equal access to all relevant information in the possession, custody, or control of the other party, the constitutional privilege against self-incrimination restricts state access to the most relevant information in criminal cases—the defendant’s own testimony—even while the “beyond a reasonable doubt” standard places a heightened burden of proof on the state to secure and present relevant evidence. The limited prosecutorial obligation to share evidence with the defendant helps to level the playing field. In addition, many jurisdictions utilize reciprocal discovery rules that take advantage of the state’s limited production obligations to enhance its ability to induce defendants to turn over information while maintaining equilibrium between the two sides (Bibas 2004, p. 2494).

This, however, gives rise to a puzzle. Notwithstanding that discovery rights in criminal litigation are limited, many prosecutors voluntarily maintain “open-file” discovery policies that allow defense attorneys ready access to the information and evidence in the prosecutor’s file even though the law does not require it (Prosser 2006, p. 593). Given limited formal discovery requirements in criminal cases in most jurisdictions, open-file discovery practices seem inconsistent with the logic of an adversarial system. So how can this apparent anomaly be explained?

Although the reasons many prosecutors support liberal informal discovery rules are undoubtedly complex (Heumann 1978), one explanation is that such policies combat defendants’ cognitive resistance to pleading guilty. By opening up the evidentiary files early in the criminal process, prosecutors can vividly demonstrate the strength of their cases to defendants. Increasing the defendant’s understanding of the evidence, and thus the accuracy of her estimate of the probability of conviction, facilitates plea bargaining by reducing the area of potential disagreement between the parties as to case values.

Informal discovery mechanisms are likely better at combating antiplea bias than formal discovery regimes. Although access to greater information allows defendants to make better estimates about trial outcomes, the effects of confirmation and self-serving bias might well limit the persuasive effect of enhanced discovery were it provided equally in all cases. Selection biases of defendants can be combated through exercise of control over the information that is shared. Informal discovery practices permit prosecutors greater discretion to choose which cases they will grant defendants liberal discovery, and to selectively withhold information in other cases simply by “neglecting” to put it into the file. Since prosecutors do not have to respond to formal motions, they risk little by such selective omission. Because the information that will be provided through open-file practices will usually be limited to inculpatory information that bolsters the strength of the prosecutor’s case, open-file policies almost certainly encourage plea bargaining by undermining defendants’ tendencies to ignore or discount information that conflicts with their preconceived views. In short, informal liberal discovery policies permit prosecutors to fully disclose where the evidence is

strong, and the discretion to withhold disclosure in weak cases where the discovery is most likely to confirm a defendant's belief that he can prevail at trial.

The cognitive impact of open-file discovery policies is further magnified in jurisdictions in which prosecutors not only open their files to defense lawyers, but also on occasion make "reverse proffers," that is, presentations of the evidence as the state would use it directly to the defendant to convince him of the strength of the case. Although reverse proffers require an added commitment of prosecutorial resources, they are an especially effective debiasing technique that facilitates a defendant's ability to overcome her tendency towards overoptimism and see the case from the prosecutor's, or the jury's, perspective.

Of course, even with perfect access to the file, calculating probabilities of conviction remains an imprecise science. No one can say for certain what will happen in any particular case should it go before a jury—the vagaries of the system are too great even for experienced prosecutors and defense lawyers to anticipate. This ambiguity should therefore lead overconfident defendants to systematically overestimate their chances at trial and to reject plea bargains even when the probabilities seem skewed heavily toward conviction. Open-file discovery policies and the strategic use of reverse proffers both help to reduce ambiguity, shift the defendant's perspective, and therefore facilitate plea bargaining.

A second feature of discovery law in the United States further enhances the potential of liberal informal discovery policies to facilitate guilty pleas. While most prosecutors willingly permit defendants access to the inculpatory evidence in the file precisely because such access facilitates guilty pleas, open access to exculpatory evidence is much more problematic. Although *Brady v. Maryland*, 373 U.S. 83 (1963), purports to require prosecutors to turn over exculpatory evidence to criminal defendants, *Brady* has been consistently construed in ways that minimize its adverse impact on plea bargaining. Of the many ways *Brady* has been limited, two stand out. First, the timing requirements imposed by *Brady* have been construed in some jurisdictions to obviate any need to produce exculpatory evidence prior to entry of a guilty plea.<sup>7</sup> As a result, many defendants will have a fresh and vivid understanding of the inculpatory evidence after having reviewed it, or heard the prosecutor present it, during plea negotiations, but no awareness of material exculpatory evidence that might be available to counterbalance the incriminating evidence. As such, the likelihood of conviction may appear greater to the defendant than is actually warranted. Second, *Brady*'s "materiality" requirement necessarily means that marginally relevant exculpatory evidence will often not be turned over to defendants, even while similarly marginal inculpatory evidence is turned over.<sup>8</sup> Thus, current discovery rules enhance the prosecutor's ability to demonstrate the strength of her case and hide its weaknesses and thereby make it easier to sell plea bargains to reluctant buyers.

<sup>7</sup> See *U.S. v. Conroy*, 567 F.3d 174 (5th Cir. 2009).

<sup>8</sup> See *United States v. Bagley*, 473 U.S. 667, 682–83 (1985) (holding that *Brady* obligations only apply if the failure to disclose exculpatory evidence passes materiality threshold).

### 3.2.2 *Minimizing Sentencing Ambiguity*

Ambiguity, of course, attends both the guilt and sentencing phases of a criminal trial. Just as relatively liberal discovery policies reduce guilt-phase ambiguity, the two most prominent developments in criminal sentencing practice—the spread of sentencing guidelines and the enactment of mandatory minimum sentences—have had similar ambiguity-reducing effects on sentencing outcomes. Defendants may not know how likely it is that they will be convicted at trial, but it is relatively easy for defense lawyers to precisely demonstrate the limited range of sentence that will accompany a conviction by referencing applicable guidelines or statutory minimums (Standen 1993). Guidelines systems and mandatory minimums counteract the tendency of defendants to believe that, if convicted, they will (or at least might) receive relatively light sentences. With guideline sentencing tables and mandatory minimum sentencing statutes in hand, lawyers can more easily demonstrate to their clients the harsh consequences of a guilty verdict at trial, and therefore convince them more easily of the relative advantages of accepting a plea bargain.

It is almost certainly not coincidental that the modern trend toward increasingly determinate sentencing schemes has paralleled the increase in guilty-plea rates (Fisher 1993). Indeed, historically, plea bargaining did not become the dominant mode of criminal adjudication until judicial discretion was checked through the rise of determinate sentencing, since the more uncertainty there is about punishment, the more difficult it is to accurately “price” a guilty plea. Modern sentencing guidelines and mandatory minimums further reduce sentencing ambiguity and thereby further facilitate guilty pleas.

Guidelines and mandatory minimum sentences also help adjust the defendant’s cognitive anchoring in ways that make plea-bargained outcomes look more desirable. Cognitive research on “anchoring” has demonstrated that, contrary to assumptions used in rational choice theory, decision-making does not solely turn on an evaluation of absolute values. Rather, decisions in the context of uncertainty are influenced by perceptions of whether an outcome represents a gain or loss measured by a perceived baseline or anchor. Because people make estimates by starting at an initial “anchor” position and adjusting it upwards or downwards, final estimates are heavily, and arbitrarily, affected by the initial anchor. Low anchors will lead to low estimates, and high anchors to high estimates.

Sentencing guidelines and mandatory minimums perform an effective anchoring function by plainly establishing high expected punishment ranges upon conviction. The threat of their application can make even onerous negotiated sentences look like gains in light of the even harsher alternatives that will necessarily apply after a trial conviction. Research on prosecutorial charging practices suggests that prosecutors in fact are more likely to charge a lesser offense against offenders who plead guilty and who would otherwise be subject to a three-strike sentence (Bjerk 2005, p. 593). Sentencing guidelines and mandatory minimums thus allow bargained-for discounts to be presented and perceived as gains from the guidelines baseline rather than as losses from the status quo. This conclusion is consistent with recent studies indicating that often

expectations—rather than the status quo—constitute the pertinent reference point in judging changes as gains or losses (Zamir 2012, p. 837).

### **3.3 Overcoming Loss Aversion and the Endowment Effect through Framing: Pretrial Detention and High Process Costs**

Loss aversion suggests that defendants will be strongly disinclined to plead guilty. A defendant's aversion to suffering a certain loss by pleading guilty, however, can be countered by framing the guilty plea as a gain rather than a loss. Although people tend to take risks to avoid losses, they are much more risk-averse when it comes to seeking gains. Although pleading guilty looks like a certain loss when the defendant is free at the time the plea is entered and in shackles the moment after, the perception of loss flowing from a guilty plea diminishes when the defendant is already behind bars.

Substantial numbers of felony defendants are detained prior to trial. Indeed, in state courts on average approximately one-third of all felony defendants are detained pretrial, and in many jurisdictions, two-thirds or more of felony arrestees typically can't make bail or are not given the option. In federal courts, approximately 72 percent of all felony defendants are detained until the completion of trial proceedings (Sourcebook of Criminal Justice Statistics 2003, pp. 414–15). Most defendants thus make the decision to plead guilty from behind bars rather than while “on the street.” The psychological effect of pretrial detention should not be underestimated. Indeed, as one experienced public defender noted, a defendant who wins pretrial release may, ironically, make worse plea-bargaining choices, because it is much more difficult for defendants who are not detained pretrial to “step in”—that is, to go from being free to being an inmate—by accepting a favorable plea bargain than for a defendant who is already locked up to accept the same plea bargain. Where the expected sentence following a guilty plea is time served, and the cost of holding out for a trial is continued detention, the perception that a guilty plea is a gain and trial a loss is virtually overwhelming.

In the United States, most criminal defendants do not pay out of pocket for their legal representation. Approximately 80 percent of criminal defendants are indigent and thus receive publicly financed legal counsel. This does not mean that indigent defendants do not pay for process—they simply pay in a different currency. Pretrial detention is just the most onerous process cost among the many burdensome process costs imposed on criminal defendants. As numerous firsthand accounts of the criminal justice system attest, even if a defendant manages to make bail or is released on his own recognizance before trial, the costs of contesting a criminal charge can be astounding. Tedious lines to get through security, strict limitations on what can be brought into the courtroom, interminable waiting for cases to be called, and seemingly endless continuances (that require working defendants and accompanying family members to burn up vacation and sick days, require childcare arrangements to be made and paid for, etc.) can make

criminal cases seem to drag on indefinitely. All of these process costs conspire to dissuade defendants from exercising their right to a trial.

These high process costs explain why almost every misdemeanor defendant, in the end, resolves his case with a guilty plea. Where “the process is the punishment” (Feeley 1979), minimizing process is the best way to minimize punishment. Pleading guilty is almost always the best route to truncating the process.

Even defendants who are incarcerated before trial must pay additional process costs to contest a criminal case. Although one might assume that a jailed defendant would have little to lose in terms of “frictional costs” by contesting a case—after all, the opportunity cost seems small at first glance—in fact even routine court appearances are often onerous for the incarcerated defendant. In many urban systems, defendants are subjected to drug-sniffing dogs that bite, strip searches, and numerous other indignities and rough treatment while waiting to appear in court. Pretrial detention coupled with the unrelenting misery endemic to most urban criminal court appearances serve an obvious functional purpose. By making the exercise of legal rights tangibly and immediately painful, high process costs reframe the decision to plead guilty.

A plea bargain presents an opportunity to cut short pretrial detention and/or to end an interminable and costly legal process, and thus what might at first look like a loss-aversion triggering event is transformed into an opportunity for a gain. Similarly, although the endowment effect suggests that a defendant charged with a crime should initially place a high value on his right to a jury trial, pretrial detention and an onerous legal process—by default or by design—undermines that assessment by turning the right to trial into a costly liability rather than an asset. So reframed, the cognitive effects of loss aversion and the endowment effect help reinforce the decision to plead guilty.

### 3.4 Correcting Overdiscounting: Lawyers’ Roles in Facilitating Guilty Pleas

A discussion of behavioral economics and plea bargaining would not be complete without consideration of the critical role played by defense lawyers in the plea-bargaining system. Defense lawyers provide defendants with more accurate estimates of the strength of the prosecutor’s case, the likely consequences of conviction, and of the “going rates” in the jurisdiction for plea bargains. Defense lawyers thus provide the most important tool necessary to permit the defendant to make a rational decision about whether to accept a plea bargain: an assessment of the value of the defendant’s trial rights. Indeed, the depiction of plea bargaining as a rational market between players making informed economic choices would be entirely implausible without the assumption of the defense bar’s role in guiding defendants to make reasonably rational decisions.



Defense lawyers may have an equally important role to play in combating the cognitive biases of their clients. By using a variety of persuasive techniques such as helping the defendant to visualize the manner in which the evidence will be used in court, defense lawyers can help to “debias” defendants and minimize cognitive errors such as overconfidence and egocentricity. For instance, one of the most important functions a defense lawyer performs is to correct their clients’ tendency to overdiscount. Some defendants discount the future at irrationally high rates, leading them to make decisions that they soon come to regret. As already discussed in this chapter, the decision to roll the dice against the odds and stand trial in the hope of winning an acquittal is in some cases the product of discounting error or bounded willpower. Even defendants who correctly understand the relative benefits of pleading guilty might nonetheless place too heavy an emphasis on their present well-being at the expense of their future well-being, while others might lack the willpower to make the utility-enhancing choice that they know will better enhance their expected long-term utility. Effective lawyering therefore frequently means changing the discounting preferences or bolstering the bounded willpower of defendants. Selling guilty pleas to clients, in other words, is one of a defense lawyer’s most important jobs. Setting aside the structural defects that plague indigent defense systems—gross underfunding and client-agent conflicts of interest to name only two (Vick 1995; Alschuler 1975)—defense lawyers contribute to the high rate of plea bargaining by correcting their clients’ discounting errors and other cognitive biases. Because of the well-documented underfinancing of the criminal defense bar, overworked and underpaid lawyers frequently have little time or inclination, and virtually no incentive, to do much else.

## 4 CONCLUSION

---

Pleading guilty to a serious criminal offense is an intuitively unappealing choice that most people can be expected to strongly resist. This intuition is supported by insights gained from behavioral economics research. Just as behavioral economics casts critical light on conventional economic models of plea bargaining that assume plea bargains represent the rational product of predictable market forces, it also provides a functional explanation for many of the most important recent developments in criminal law, as well as some of the most unattractive features of typical urban criminal justice systems. The need to overcome criminal defendants’ cognitive resistance to pleading guilty helps explain not only why determinate sentencing and mandatory minimums have proved so popular among courts, prosecutors, and politicians, it also explains the prevalence of pretrial detention, the harsh rigors of courthouse routines, the long lines, endless waits, and strict conduct rules in misdemeanor courts, and the generous grants of continuances to criminal litigants. While many of the most important developments in modern criminal law, especially in the United States—the inflation of sentences, the expansion of prison populations, the advent of sentencing guideline

regimes, and the widespread adoption of mandatory minimum and career criminal sentencing provisions—undoubtedly are the product of a complex of political, cultural, historical, and administrative causes, the uniformity with which they contribute to the cause of inducing guilty pleas by overcoming the cognitive resistance of defendants cannot be overlooked. Further research is needed, however, to better understand how plea-bargaining decisions are affected by the wide array of discoveries made by cognitive researchers, and how the interplay of these various influences on rational decision-making affects the integrity and fairness of outcomes.

By viewing the criminal justice system as an integrated plea-bargaining machine that functionally works to overcome the cognitive resistance of criminal defendants to plead guilty, we gain important perspective on the system's components. This perspective strongly suggests that isolated reforms of various criminal justice processes are not likely to succeed, at least not if their goal is to improve defendant decision-making or systemic sorting accuracy. Rather, what is needed is system-wide transformation. Recognition of the “cognitive design” of the criminal justice system should cause us to question some of the most fundamental notions regarding the function of courts and of legal process in the administration of punishment. Meaningful reform of the criminal justice system is unlikely to occur absent an abandonment, or at least a reduction, of the drive to increase systemic leverage to obtain guilty pleas while minimizing process costs. Indeed, the very notion of maximizing “efficiency” in the prosecution of crime might need to be reevaluated—a project that takes us well beyond the domain of both rational choice theory and behavioral economics.

## REFERENCES

- Alschuler, A. 1975. The defense attorney's role in plea bargaining. *Yale Law Journal* 84:1179–315.
- Alschuler, A. 1981. The changing plea bargaining debate. *California Law Review* 69:652–730.
- Babcock, L., and G. Loewenstein. 2000. Explaining bargaining impasse: The role of self-serving bias. In *Behavioral Law and Economics*, C. R. Sunstein. New York: Cambridge University Press, 109–26.
- Bibas, S. 2004. Plea bargaining outside the shadow of trial. *Harvard Law Review* 117:2464–547.
- Birke, R. 1999. Reconciling loss aversion and guilty pleas. *Utah Law Review* 1999:205–54.
- Bjerk, D. 2005. Making the crime fit the penalty: The role of prosecutorial discretion under mandatory minimum sentencing. *Journal of Law and Economics* 48:591–626.
- Boari, N., and G. Fiorentini. 2001. An economic analysis of plea bargaining: The incentives of the parties in a mixed penal system. *International Review of Law and Economics* 21:213–31.
- Bureau of Justice Statistics. 2002. Available at <http://www.ojp.usdoj.gov/bjs/pub/pdf/scscfo2.pdf>.
- Camerer, C., and R. Thaler. 1995. Anomalies: Ultimatums, dictators and manners. *Journal of Economic Perspectives* 9:209–19.
- Covey, R. 2007. Reconsidering the Relationship between Cognitive Psychology and Plea Bargaining. *Marquette Law Review* 91:213–47.
- Covey, R. 2013. Police misconduct as a cause of wrongful convictions. *Washington University Law Journal* 90:1133–89.

- Department of Justice, Bureau of Justice Statistics. Sourcebook of Criminal Justice Statistics Online. Table 5.22.2009. <http://www.albany.edu/sourcebook/pdf/t5222009.pdf>.
- Feeley, M. 1979. *The Process Is the Punishment: Handling Cases in a Lower Criminal Court*. New York: Russell Sage Foundation.
- Fisher, G. 1993. *Plea Bargaining's Triumph: A History of Plea Bargaining in America*. Stanford, CA: Stanford University Press.
- Gazal-Ayal, O., and A. Tor. 2011. The innocence effect. *Duke Law Journal* 62:339–401.
- Gigerenzer, G. 2012. Moral satisficing: Rethinking moral behavior as bounded rationality. *Topics in Cognitive Science* 2:528–54.
- Glockner, A., and C. Engel. 2012. Behavioral decision making: Role induced bias in court: an analysis. First published online, April 11. doi:10.1002/bdm.1761. <http://onlinelibrary.wiley.com/doi/10.1002/bdm.1761/pdf>.
- Heumann, M. 1978. *Plea Bargaining: The experiences of Prosecutors, Judges, and Defense Attorneys*. Chicago: University of Chicago Press.
- Jolls, C., and C. R. Sunstein. 2006. Debiasing through law. *Journal of Legal Studies* 35:199–242.
- Jolls, C., C. R. Sunstein, and R. Thaler. 1998. A behavioral approach to law and economics. *Stanford Law Review* 50:1471–550.
- Kahneman, D., J. Knetsch, and R. Thaler. 1991. The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives* 5:193–206.
- Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47:263–91.
- King, N., D. Soule, S. Steen, and R. Weidner. 2005. When process affects punishment: Differences in sentences after guilty plea, bench trial, and jury trial in five guidelines states. *Columbia Law Review* 105:959–1009.
- Korobkin, R., and T. Ulen. 2000. Law and behavioral science: Removing the rationality assumption from law and economics. *California Law Review* 88:1051–144.
- Kunen, J. 1983. *How Can You Defend Those People? The Making of a Criminal Lawyer*. New York: Random House.
- Langer, M. 2004. From legal transplants to legal translations: The globalization of plea bargaining and the Americanization thesis in criminal procedure. *Harvard International Law Journal* 45:1–64.
- Luna, E., and M. Wade. 2010. Prosecutors as judges. *Washington & Lee Law Review* 67:1413–532.
- McCoy, C. 2003. Bargaining in the shadow of the hammer: The trial penalty in the USA. In *The Jury Trial In Criminal Justice*, D. D. Koski. Durham, NC: Carolina Academic Press, 23–9.
- Mnookin, R., and L. Kornhauser. 1979. Bargaining in the shadow of the law: The case of divorce. *Yale Law Journal* 88:950–97.
- Prosser, M. 2006. Reforming criminal discovery: Why old objections must yield to new realities. *Wisconsin Law Review* 2006:541–614.
- Rachlinski, J. 2003. The uncertain psychological case for paternalism. *Northwestern University Law Review* 97:1165–226.
- Reinganum, J. 1988. Plea bargaining and prosecutorial discretion. *American Economics Review* 78:714–28.
- Roth, A. 1995. Bargaining Experiments. In *The Handbook of Experimental Economics*, J. Kagel and A. Roth. Princeton, NJ: Princeton University Press, 253–348.
- Simon, H. 1992. Introduction. In *Economics, Bounded Rationality and the Cognitive Revolution*, M. Egidi and R. Marris. Brookfield, VT: Edward Elgar, vii–viii.

- Sourcebook of Criminal Justice Statistics. 2003. Available at <http://www.albany.edu/sourcebook/pdf/t513.pdf>.
- Standen, J. 1993. Plea bargaining in the shadow of the guidelines. *California Law Review* 81:1471–540.
- Teichman, D. 2011. The optimism bias of the behavioral analysis of crime control. *University of Illinois Law Review* 2011:1697–712.
- Vick, D. 1995. Poorhouse justice: Underfunded indigent defense services and arbitrary death sentences. *Buffalo Law Review* 43:329–460.
- Weinstein, N. 1980. Unrealistic optimism about future life events. *Journal of Personality and Social Psychology* 39:806–20.
- Wright, R., and R. Engen. 2007. Charge movement and theories of prosecutors. *Marquette Law Review* 91:9–38.
- Zamir, E. 2012. Loss aversion and the law. *Vanderbilt Law Review* 65:829–94.
- Zeisel, H. 1980. The offer that cannot be refused. In *The Criminal Justice System: Materials on the Administration and Reform of the Criminal Law*, F. Zimring and R. Frase. Boston: Little, Brown, 558–61.

## CHAPTER 26

---

# JUDICIAL DECISION-MAKING

## *A Behavioral Perspective*

---

DORON TEICHMAN AND EYAL ZAMIR

### 1 INTRODUCTION

---

DRAWING on psychological insights to explain judicial decision-making goes back more than eighty years (Frank 1930). However, most of the progress in the application of insights and methodologies of judgment-and-decision-making research to judicial decision-making has been made in the past two decades. This chapter focuses on these recent studies.

The judicial system and the behavioral sciences interface at manifold points. It is thus useful to begin our review by differentiating the behavioral perspective from other bodies of research. To begin with, behavioral analysis of judicial decision-making should be distinguished from forensic psychology, that is, the contribution of psychologists to the operation of the court system through the provision of expert testimonies in legal proceedings (see generally Cutler and Kovera 2011). Judicial decision-making is also to be distinguished from judges' verbal and nonverbal communicative behavior in court (Blanck et al. 1990).

Behavioral analysis of judicial decision-making should likewise be differentiated from the empirical study of the relationships between judges' decisions, their ideological inclinations, and law as a system of norms and an institution. This latter body of research, connected in part to rational choice theory, is primarily the province of political scientists. It focuses on higher court decisions regarding public and constitutional issues. The following quote nicely summarizes the different theories:

What are the considerations that drive justices? The answer depends on who you ask, and when. Legal model researchers would reply that justices' considerations are rooted in the law, and that the law is essentially comprised of its plain meaning, the lawmaker's intent, and precedent. *Attitudinal model* researchers would

answer that the law camouflages the true main consideration: the judge's attitude vis-à-vis the facts of the case. A *rational choice model* supporter would agree with the attitudinalists, yet add strategic considerations related to the odds of the judge's opinions being accepted and ultimately realized. A *neo-institutionalist* would also claim that judges' own attitudes are a main consideration. Such a researcher would also agree that strategic considerations, which often stem from various institutional arrangements and norms, influence judges' decisions. Yet in the eyes of the neo-institutionalist, institutional variables also have their own influence on court decision making. A main institutional variable is "the law," which is defined as a dynamic mindset instead of a formal definition as per the legal model. (Weinshall-Margel 2011, p. 569, emphases added)<sup>1</sup>

Notwithstanding their disparities, these camps share some basic assumptions, including the belief that judges' decisions are driven only by their goals, the primary goal being to make good legal policy (whatever role "the law" plays in forming such a policy). These theories hardly take into account insights from cognitive and social psychology that cast doubt on these assumptions (Baum 2010; Martinek 2010). Furthermore, they are less relevant to the great majority of run-of-the-mill judicial decisions made by lower courts. Surveying this literature would thus exceed the boundaries of the present chapter.

Since judges are generally insulated from market incentives and since their decisions in particular cases do not directly affect their own well-being, standard economic analysis—which assumes that people are rational maximizers of their own utility—is not very helpful in explaining judicial behavior. In an attempt to meet this challenge, Richard Posner (1993) suggested drawing an analogy between judges' decisions and those made by managers of nonprofit enterprises (whose income does not depend on the profits of their enterprise), people who vote in political elections (despite the infinitesimal probability that their vote would affect election outcomes), and theater spectators (who identify with the characters and form an opinion on their entitlements). Contrary to the attitudinal and rational choice models, Posner portrays judges as driven by a multitude of motives going beyond making good legal policy, including a preference for expending less effort, yearnings for prestige and popularity, aversion to being reversed by higher courts, and the desire to move the docket (see also Epstein, Landes, and Posner 2013; Choi, Gulati, and Posner 2012).

Though important and insightful, Posner's more complex depiction of judicial decision-making still disregards the behavioral perspective. It does not overcome the basic difficulty facing economic analysis, as it largely assumes that judges derive utility from "playing by the rules" of judicial decision-making, advancing the public interest, and so forth. While this analysis may explain why judges do not decide randomly, or why

<sup>1</sup> On the attitudinal model, see also Segal and Spaeth (1993, 2002); on strategic models: Epstein and Knight (1998); and on neo-institutionalism: Clayton and Gillman 1999; Richards and Kritzer 2002; Bailey and Maltzman 2011.

they exert effort in the absence of direct monetary incentives, it sheds little light on how judges make decisions (see also Siegel 1999).

Hence, unlike other spheres, in which behavioral law and economics has largely evolved as a reaction to standard economic analysis, in the absence of established economic analysis of judicial decision-making, the behavioral studies in this sphere do not relate to the economic perspective. In conformity with the general spirit of this volume, we nevertheless focus on those aspects of the behavioral research that are more closely connected to the main themes of behavioral law and economics. It would therefore be useful to describe the contribution of the behavioral studies of judicial decision-making by referring to the three primary contrasts between the economic and the behavioral perspectives described in the opening chapters of this volume. These are deviations from the assumption of cognitive rationality (the heuristics and biases literature), departures from the assumption of motivational rationality (other-regarding preferences), and divergences from consequentialist morality (see chapter 1 by Baron, chapter 2 by Gächter, and chapter 3 by Baron in this volume, respectively).

Most behavioral studies of judicial decision-making belong to the heuristics and biases school of research. Hence, these studies constitute the lion's share of this chapter. The chapter also discusses the contributions of social psychology to understanding of decision-making in small groups, such as juries. Since almost no one argues that judges and juries make their decisions with a view to directly maximizing their own utility, the behavioral studies of human motivation are less relevant here. Finally, legal decision-making inevitably involves normative deliberation. In this respect, psychological studies of people's normative convictions or "moral heuristics" are important for understanding judicial decisions. However, since these issues are discussed elsewhere in this volume (chapter 11 by Zamir), they are not discussed here.

To keep the discussion manageable, we exclude a systematic analysis of the vast literature on jury instructions (see, e.g., Eisenberg and Wells 1993; Lieberman and Sales 1997; Vidmar and Hans 2007, pp. 158–68, 175–76, 236–40, 260–62). For the same reason, we only describe some of the contribution of behavioral studies to the large body of literature on the interactions between race and judicial decision-making (see, e.g., Mustard 2001; Mitchell et al. 2005).

A final note regarding the scope of this chapter is pertinent. To fully understand judges' and juries' decision-making, one must pay heed to the psychology of other key figures in the adjudication process: litigants, attorneys, and witnesses. However, delving into these issues would dramatically expand the scope of this chapter; we therefore exclude them as well. For an overview of behavioral analyses of litigation and settlement, see chapter 24 by Robbennolt in this volume; for a recent synopsis of the literature on eyewitness identification see Brewer and Wells (2011); and for a critical review of the behavioral analysis of evidence law, see chapter 27 by Vars in this volume.

This chapter thus dwells on studies highlighting the susceptibility of judicial decision-makers to the various heuristics and biases identified by cognitive psychology. The chapter is structured as follows. Section 2 presents general theories of the cognitive process of judicial decision-making. Sections 3 through 7 describe a series



of well-known cognitive phenomena and their reflection in judicial decision-making. These include the compromise and contrast effects, the effect of legally irrelevant information, the hindsight bias, the omission bias and related phenomena, and the role of anchoring in converting qualitative into quantitative judgments. Section 8 then examines fact-finders' reluctance to impose liability based on certain types of evidence. Section 9 describes the contribution of behavioral studies to better understanding judicial prejudice. Section 10 describes experimental studies of quintessentially judicial decisions: the application of legal norms to facts, and specifically the effect of the choice between rules and standards on the predictability of judgments. Sections 11 and 12 provide an overview of two fundamental questions in the behavioral analysis of judicial decision-making, discussed throughout this chapter: group decision-making and judges' versus laypersons' decision-making. Section 13 offers a general assessment of the behavioral research of judicial decision-making.

## 2 THE STORY MODEL AND COHERENCE-BASED REASONING

---

Before turning to reflect on specific heuristics and biases in judicial decisions, this section briefly presents some general theories of the cognitive mechanisms through which judicial decision-makers process complex information and reach their decisions. These theories set the framework for understanding cognitive phenomena such as priming in adjudication, disinclination to base liability on statistical evidence, and differentiating between quantitative and qualitative judicial decisions.

Primary contributions to this body of research include the work of Nancy Pennington and Reid Hastie on the story model (see Pennington and Hastie 1991), and the studies of Keith Holyoak, Stephen Read, and Dan Simon on constraint satisfaction and coherence-based reasoning (see Simon 1998, 2004). While the story model focuses on fact-finding, studies of coherence-based reasoning refer to decisions on legal issues as well. The latter studies rest on connectionist models of mental representations (for an overview see Simon 2004; Robbennolt, Darley, and MacCoun 2003; Read and Simon 2012).

The story model contests previous theories of fact-finding, such as Bayesian probability theory, algebraic models that attribute differential weights to pieces of evidence, and stochastic process models (Hastie 1993). Based on interviews and experimental studies (Pennington and Hastie 1986, 1988, 1992), the story model appears to better describe the actual psychological process of fact-finding in adjudication. According to the story model, story construction—the creation of a narrative that explains the diverse items of evidence that have been deemed reliable and relevant—is the core cognitive process determining the facts in adjudication relevant. Interviews and experiments have indicated that the mental representation of the evidence is not structured

according to factors such as the order of its presentation in court, the pertinent legal issues, or whether the evidence supports or undermines the plaintiff's version of the events. Rather, fact-finders structure the evidence to create a story. The story is constructed from three types of knowledge: the evidence presented at trial, knowledge about events similar to the one in dispute, and general notions of what constitutes a complete story. Story constructing is an active process, resulting in one or more interpretations of the evidence. When faced with different interpretations, fact-finders adopt the one that best explains the evidence, that is, the story that is most coherent and provides broadest coverage of the evidence. Coherence requires that the story contains no internal contradictions or missing elements, and that it conforms to fact-finders' beliefs about the physical world and people's motivations and behavior.

Complex stories often comprise several episodes. They include motivations, actions, and consequences, connected by physical and intentional causality. Some of the events—or their elements—may not be supported by any direct evidence and require the drawing of inferences (Pennington and Hastie 1991). Hence, the same set of evidence often gives rise to more than one story, with different fact-finders finding different stories more or less compelling. Pennington and Hastie (1988, 1991, 1992) found that the more complete and coherent a story is, and the more it covers the available evidence, the more confident are the fact-finders about its accuracy. Fact-finders' confidence about a story's veracity is further enhanced by its uniqueness, that is, by the lack of plausible alternative stories that could account for the evidence.

Pennington and Hastie (1988) also found that when the evidence supporting a possible conclusion was presented in a chronological and causal story order, while the evidence supporting rival conclusion presented in a nonstory order (e.g., according to the order of the testimonies), subjects tended to adopt the former conclusion. Even more intriguingly, when, following hearing of the evidence, subjects were presented with sentences allegedly describing the evidence, including lure sentences referring to facts that had not been included in the evidence, they were almost twice as likely to recognize lure sentences supporting their adopted story as they were to recognize lure sentences supporting the alternative one (Pennington and Hastie 1988). Fact-finders thus use different techniques to fill in gaps and strengthen their story.

The idea that fact-finding involves a choice between possible narratives is supported by an experimental study showing that evidence judged as only weakly supporting one side's version increased fact-finders' confidence in the truthfulness of the opposite version (McKenzie, Lee, and Chen 2002).

According to the story model and coherence-based theories, while the chosen story and legal conclusions may change during the trial and even following the jury instructions, the coherent story and its legal implications are not a post hoc justification of the decision. They are created during the trial and the decision-making process (Pennington and Hastie 1991; Holyoak and Simon 1999; Simon et al. 2001). This observation is closely connected to a central tenet of coherence-based theories of judicial decision-making, namely that the decision process is bidirectional (Holyoak and Simon 1999). The strength of evidence and arguments not only determines which

story the decision-maker will adopt and what decision she will make. The adopted story and decision concomitantly determine the assessed relevance, reliability, and importance of different pieces of evidence as well as the power of competing legal arguments. Subjects tend to attribute greater weight to evidence items and legal arguments that support their decision, and lesser weight to evidence and arguments opposing it (Pennington and Hastie 1988; Holyoak and Simon 1999; Simon, Snow, and Read 2004). As a corollary, a specific piece of evidence, or argument, can indirectly influence the assessed reliability or persuasiveness of other pieces of evidence and legal arguments, even absent any plausible relationship between the two.

Even if the evidence and legal arguments are initially confusing and incoherent, this bidirectional process tends to yield a conclusion that decision-makers sincerely believe to be clear and conclusive. Decision-makers are typically unaware of this coherence shift as they do not accurately recall their original assessment of the evidence and the legal argumentation (Holyoak and Simon 1999; Simon 2004). Thus, when Simon, Snow, and Read (2004) presented their subjects with a new piece of evidence that was sufficiently powerful to cause some of them to switch their initial verdict, the final verdict was accompanied by a corresponding (second) coherence shift; and switchers were no less confident in their final verdict than were those who did not switch.

The story model and coherence-based theories of judicial decision-making arguably have normative implications. Simon (2004), for instance, argued that to enhance the accuracy of juries' fact-finding, instructions about substantive law should be given prior to hearing the evidence, because subsequent instructions are unlikely to alter a coherent story formed on the basis of inaccurate assumptions about the law. Another possible implication pertains to the admissibility of prejudicial evidence. Given that a sufficiently strong piece of evidence can affect the entire mental model of the case by indirectly influencing other variables, the admission of prejudicial evidence may be more detrimental than assumed, for example, by a Bayesian theory of fact-finding, because it affects the assessed reliability and relevance of pieces of evidence regarding substantively unrelated issues (Simon 2004).

After this introduction to general theories of judicial decision-making, we turn to a discussion of specific cognitive phenomena.

### 3 COMPROMISE AND CONTRAST EFFECTS

---

Rational choice theory assumes that the relative ranking of two options is context-independent. That is, the ranking of two options should not be influenced by the sheer availability of additional options. To illustrate, a customer in a restaurant should not change her ranking of the steak and chicken options simply because a fish platter is added to the menu. And yet empirical findings from the area of consumer choice have demonstrated that decisions are often context-dependent. Adding an additional, even irrelevant, option to the mix can in fact alter people's decisions.

Researchers have identified two primary mechanisms of context dependence: the compromise effect and the contrast effect.

The compromise effect alludes to people's tendency to choose intermediate rather than extreme options. For example, when consumers were asked to choose between a medium-priced and a low-priced camera, each type was chosen by 50% of the participants. However, when asked to choose between those two cameras and an additional high-end camera, 72% chose the medium option (Simonson and Tversky 1992). Kelman and his colleagues (1996) demonstrated that the compromise effect can influence judicial decisions. Their experiment focused on conviction decisions with respect to different types of homicide offenses: manslaughter, murder, and murder with aggravating circumstances. The results showed that much like in the case of cameras, the introduction of an additional more-severe offense pulled fact-finders towards the intermediate option. Facing a choice between manslaughter and murder, 47% of the subjects chose manslaughter while 53% chose murder. When the third option was added, only 19% of the subjects chose the manslaughter option whereas 39% chose murder and 42% chose murder with aggravating circumstances.

Another type of context dependence is the contrast effect. Adding an option that highlights the attributes of one of the products being evaluated can cause people to choose that product even though the added option itself is strictly inferior and therefore irrelevant to the decision. For instance, when facing a choice between a Cross pen and \$6 in cash, only 36% of the subjects chose the pen. When a third irrelevant option was introduced: a pen that is clearly inferior to the Cross pen, 46% of the subjects preferred the Cross pen over the money. Apparently, the inferior pen altered the way in which people attached value to the Cross pen (Simonson and Tversky 1992). Kelman and his colleagues (1996) showed that the contrast effect can also influence legal choices. Participants in their study were asked to choose the sanction suitable for a criminal. When deliberating between jail and probation, the introduction of an inferior sanctioning option that highlighted the advantages of the probation option caused more people to choose it. Similar results were reported by Rachlinski and Jourden (2003) with respect to years of sentencing.

More work remains to be done before we acquire a good understanding of the way in which context dependence influences judicial decision-making. First, most studies have thus far focused on laypersons' decisions (for an exception see Rachlinski, Wistrich, and Guthrie 2013). Arguably, decision-makers more familiar with the legal decision environment will be less prone to influences from irrelevant factors. Second, the handful of existing studies do not account for the many nuances associated with different legal questions. For example, while Rachlinski and Jourden (2003) identified a contrast effect with respect to years of imprisonment, they did not find such an effect with respect to the death penalty: "Death, it seems, is different after all" (Rachlinski and Jourden 2003, p. 482).

## 4 IRRELEVANT INFORMATION

---

The courtroom provides a unique decision-making environment. Whereas the human mind is trained to incorporate all available information so as to render the best decision, this is not always the case in court. Both the rules of evidence and substantive legal rules sometimes dictate the exclusion of certain facts from the information set available to the decision-maker. In addition, attorneys might willfully introduce legally irrelevant information in order to influence decisions. In this section we examine the degree to which such information affects judicial decision-makers.

To begin with, exclusionary rules are premised on the assumption that the prejudicial effects of certain types of evidence outweigh their probative value. For example, information regarding the defendant's past convictions might be relevant for the determination of liability in the case at hand but it may also skew decisions towards a finding of guilt. Other types of evidence, such as hearsay testimony, can be excluded due to their limited probative value. Finally, some exclusionary rules stem from policy considerations unrelated to the probative weight of the evidence. For example, evidence obtained through illegal police practices might be deemed inadmissible in order to incentivize police to behave appropriately in future cases and to protect the fairness of the judicial process.

During adjudication, however, fact-finders are often exposed to evidence that is later determined to be inadmissible. This can occur when a witness exposes the inadmissible evidence in the courtroom. Inadmissibility can also characterize information coming from external sources, such as the media. The question of whether fact-finders actually manage to ignore inadmissible information has long troubled legal scholars and courts. In the past couple of decades scholars have turned to examine this question empirically.

A case in point is the influence of information about the defendant's past convictions. Taking an experimental approach, Greene and Dodge (1995) exposed one group of mock jurors to the defendant's past convictions while the other group of mock jurors was not exposed to such information. Those mock jurors informed about the past convictions were significantly more likely to reach a guilty verdict (see also Wissler and Saks 1985; Tanford and Cox 1988). More recently, Eisenberg and Hans (2009) took an observational approach to the topic. As part of their study, they assembled a unique data set documenting the behavior of defendants in criminal trials. As Eisenberg and Hans show, criminal records are often introduced into evidence when the defendant takes the stand. When introduced at that stage of the trial, such records induced a significant rise in conviction rates in close cases, that is, cases where the evidence presented by the prosecution is not overwhelmingly strong. Notably, this increase is not driven by the effect of the past conviction on the defendant's credibility (the reason for which past convictions are usually admitted). Rather, it is driven by the effect it has on the jury's decision threshold.

Numerous studies have documented the effects of inadmissible evidence in other legal domains, such as hearsay evidence (e.g., Schuller 1995), pretrial media reports (e.g., Fein, McCloskey, and Tomlinson 1997), and illegally obtained evidence (e.g., Kerwin and Shaffer 1994). These studies show that inadmissible evidence affects judicial decision-making in civil as well as criminal settings, irrespective of whether that evidence favors the prosecution or the defense.<sup>2</sup> A recent meta-analysis concluded that “[i]nadmissible evidence produced a significant impact on guilty verdicts” (Stebly et al. 2006, p. 477). While this impact is relatively small, it is statistically significant (Stebly et al. 2006, p. 486).

Researchers have also documented similar behavior among professional judges. In a study focusing on inadmissible evidence regarding remedial measures in a product liability case, Landsman and Rakos (1994) found that both judges and mock jurors were unable to disregard the facts they were required to disregard. The authors concluded that in the setting examined, “judges and jurors may not be very different in their reactions to potentially biasing material” (Landsman and Rakos 1994, p. 125). In a later series of experiments, Wistrich, Guthrie, and Rachlinski (2005) reported somewhat more nuanced results. They found that judges could not ignore inadmissible evidence regarding settlement offers, privileged information, prior sexual history of a rape victim, prior criminal records of a plaintiff, and information that the government had agreed not to use in trial. At the same time, judges did manage to ignore a confession that was obtained in violation of a defendant’s right to counsel and the outcome of a search that was conducted without probable cause. The authors carefully acknowledge that the pattern of results they observed “defies easy explanation” and requires more data. Clearly, they are correct in their assessment of their results.<sup>3</sup>

Finally, a question of significant practical importance is the extent to which admonitions made by courts to disregard inadmissible evidence do alter jurors’ decisions. Since the “human mind cannot simply forget information on command,” there is room for some skepticism as to the effectiveness of such instructions (Daftary-Kapur, Dumas, and Penrod 2010, p. 138). Indeed, the mentioned meta-analysis suggests that “judicial instruction did not return verdicts to the level generated by jurors never exposed to the inadmissible evidence” (Stebly et al. 2006, p. 478).

Aside from this general finding, two other findings related to admonitions merit closer attention. First, admonitions may draw greater attention to the inadmissible evidence in some cases and thus produce a backlash. Pickel (1995), for example, reported that mock jurors who were exposed to a detailed admonition exhibited a greater

<sup>2</sup> That said, the vast majority of studies deal with incriminating evidence in criminal cases (Stebly et al. 2006, p. 476).

<sup>3</sup> While the inability (or limited ability) to disregard information challenges the effectiveness of exclusionary rules, this inability may actually justify such rules. Guttel (2004) argued that unreliable evidence—evidence that is likely to be refuted by the other party, such as hearsay—should be excluded altogether, because its subsequent refutation may distort fact-finders’ decisions due to the cognitive phenomenon of overcorrection.

tendency to convict when compared to mock jurors who were only informed about the inadmissibility of the evidence (see also Cox and Tanford 1989; Lee, Krauss, and Lieberman 2005). Second, jurors are sensitive to the admonition's content. When jurors are required to disregard a certain piece of evidence due to a technicality, they are relatively reluctant to do so. In contrast, when jurors are asked to ignore evidence due to its limited probative weight, they exhibit a greater tendency to comply with instructions (Stebly et al. 2006, p. 487).

## 5 HINDSIGHT BIAS

---

Courts are frequently called upon to evaluate a decision in retrospect, after the decision's outcomes are known. In negligence cases, for example, the reasonableness of the precautions taken by the defendant are examined subsequent to materialization of the risk associated with a harm. Yet the law often requires judicial decision-makers to ignore the outcome information revealed *ex post* and evaluate the issues from a purely *ex ante* perspective. Behavioral findings suggest, however, that decision-makers find it difficult to ignore such information.

Starting with Fischhoff (1975), a large body of work has documented the existence of hindsight bias (for a review and meta-analysis see Guilbault et al. 2004; Roese and Vohs 2012). Studies have shown that people tend to overestimate the probability of an event once they are aware of the fact that it has occurred. Researchers have provided cognitive as well as motivational explanations for the bias (Guilbault et al. 2004).

Hindsight bias has been examined in numerous legal contexts (see chapter 14 by Teichman in this volume), including determination of negligence in tort cases (Kamin and Rachlinski 1995), analysis of novelty in patent law (Mandel 2006), and the finding of probable cause in search and seizure cases (Rachlinski, Guthrie, and Wistrich 2011). Studies have also demonstrated that professional judges are susceptible to such bias (Guthrie, Rachlinski, and Wistrich 2001) although they may deal with it more appropriately than do laypersons (Hastie and Viscusi 1998). Finally, debiasing decision-makers from the effects of hindsight bias has proven to be a thorny task (Kamin and Rachlinski 1995).

## 6 OMISSION BIAS AND RELATED PHENOMENA

---

Omission bias is the tendency to prefer inaction to action when facing risky alternatives. People are considered to bear greater moral responsibility for harmful outcomes they actively brought about than for those they brought about passively (Spranca,



Minsk, and Baron 1991). People anticipate experiencing greater regret if their action would result in worse outcomes than inaction, compared to the regret they expect to feel if they would refrain from action and it turned out that action would have produced better outcomes (Kahneman and Tversky 1982a). People might therefore prefer harmful omissions to less harmful commissions (Ritov and Baron 1990).

Presumably, a judge called to decide a case cannot refrain from delivering a judgment; hence omission bias may seem irrelevant to judicial decision-making. However, there is experimental support for the claim that laypersons perceive accepting a claim as more active than dismissing it (Zamir and Ritov 2012). If this perception is shared by judges, omission bias may help explain why they are reluctant to accept claims even when the plaintiff's version of the facts is slightly more persuasive than that of the defendant. The general standard of proof in civil litigation in common-law systems is preponderance of the evidence. Plaintiffs have to establish their case as more probable than not in order to prevail. Notwithstanding this formal rule, in a series of experiments conducted with advanced-years law students and experienced lawyers, Zamir and Ritov (2012) found that the actual standard of proof is considerably higher. To accept a claim, the decision-maker should rate the persuasiveness of the plaintiff's version around 70 on a scale of 0 to 100 (where 0 indicates that there is no doubt that the plaintiff's version is incorrect and 100 indicates that there is no doubt that it is correct). These experiments provide a *prima facie* evidence that judges exhibit an omission bias. Guthrie and George (2005) have similarly suggested that omission bias can explain the strong tendency of appellate courts to dismiss appeals.

Psychologists have also identified a closely related phenomenon, *status quo bias*. Other things being equal, people tend to stick to the state of affairs they perceive as the *status quo* rather than opt for an alternative (Samuelson and Zeckhauser 1988; Kahneman, Knetsch, and Thaler 1991). Ordinarily, changing the *status quo* requires some action, whereas retaining the *status quo* involves mere omission. Hence, the *status quo bias* is usually confounded with the omission bias although they can exist separately (Schweitzer 1994) and can pull in opposite directions (Ritov and Baron 1992).

*Status quo bias* has been mentioned as a possible explanation for court reluctance to issue preliminary injunctions that disrupt the *status quo* (Zamir 2012) as well as to appellate court aversion to reversing lower court decisions (Guthrie and George 2005). It has also been proposed as a possible explanation for court adherence to the doctrine of *stare decisis* (Prentice and Koehler 2003, p. 638; Jois 2009), and more generally to the great influence that the past exerts on current law (Wistrich 2012). And yet, when Zamir and Ritov (2012) presented their subjects with a scenario in which dismissing a claim for a declaratory judgment would alter the *status quo*, while accepting it would sustain the *status quo*, the omission bias seems to have had a greater impact on the decision.

Another related phenomenon is escalation of commitment. Expected utility theory posits that in choosing between different courses of action, only future costs and benefits should be taken into account because the past cannot be changed. This implies that unrecoverable, incurred costs that will not affect future costs or benefits should not

influence current decisions. However, numerous laboratory and field experiments as well as empirical studies have established that people very often do consider sunk costs when making decisions. People thus tend to continue endeavors the more resources, time, or efforts have already been invested in them (Arkes and Blumer 1985; Staw and Ross 1989).

Gely (1998) has suggested that the rich literature on escalation of commitment and its psychological, social, and institutional determinants can fruitfully contribute to better understanding the application of the concept of *stare decisis*, the binding force of precedents that is a cornerstone of the common law. There is much room for future research of judicial decision-making and the sunk costs effect. For example, under the doctrine of mootness, a court should halt adjudication and dismiss the case once the dispute has become academic because, for instance, the defendant agency has abandoned the policy challenged by the petitioner. It would be interesting to examine whether court receptiveness to mootness claims might depend on the amount of judicial resources already spent on the case.

## 7 CONVERTING QUALITATIVE INTO QUANTITATIVE JUDGMENTS AND THE ANCHORING EFFECT

---

### 7.1 Legal Numbers

A large body of literature is devoted to judicial decisions involving quantification, such as the award of monetary damages and the determination of criminal fines and length of imprisonment. Such numerical decisions are often described as variable and unpredictable. Some of this variability and unpredictability may be due to people's limited proficiency with numbers, especially among lay jurors (Hans and Reyna 2011). Other sources of variability that are again likely to be more glaringly reflected in lay jurors' decisions are related to misunderstandings regarding the relevant legal criteria, or with discrepancies between formal legal rules and people's normative judgments. For example, when determining damages for a plaintiff's economic losses, judicial decision-makers may take the defendant's wealth, malice, or gains from violating the plaintiff's rights into account even if these factors are considered legally irrelevant.

These causes of variability apply to numerical decisions irrespective of whether conversion from one scale to another is required. In awarding monetary damages for economic losses, for example, both scales are monetary. When numerical judgments require conversion from one scale to another, such as conversion of the defendant's culpability, into the length of imprisonment, difficulties can multiply. Much of the research centers on this particular difficulty.

While conversion from one scale to another inherently poses a challenge, the scope of this challenge can vary according to the respective legal norms and decision-maker expertise. At one extreme we find cases where the law allows largely unfettered discretion, for example, when setting “reasonable” damages for pain and suffering. Such unfettered discretion is particularly troublesome when decision-makers such as lay jurors lack the relevant experience and meaningful reference points. This situation persists in many jurisdictions in the United States (Chase 1995; Greene and Bornstein 2003, pp. 175–76). At the other pole we find no room for discretion, as when the law lays down precise sanctions or remedies. Between these two extremes, more or less specific guidelines for quantification can be set, as some legal systems have done regarding criminal sentencing.

This section surveys behavioral studies of quantitative judicial decision-making, focusing on the conversion of qualitative into quantitative judgments.

## 7.2 Qualitative and Quantitative Judgments: Empirical and Experimental Findings

When legal decision-makers convert qualitative judgments into quantitative ones—as in the award of damages for nonpecuniary harms and the imposition of criminal sanctions—their decisions should ideally be consistent, predictable, and justifiable. The numbers should serve the law’s goals, such as just desert and deterrence. They should be predictable *ex ante* and justifiable *ex post* (Kahneman, Schkade, and Sunstein 1998; Hastie 2011).

Empirical and experimental studies reveal a remarkable degree of similarity and predictability in the qualitative judgments made by judges and jurors regarding issues such as the severity of the plaintiff’s injury (Wissler, Hart, and Saks 1999), the outrageousness of a defendant’s behavior, and the appropriate severity of punishment (Kahneman, Schkade, and Sunstein 1998). At the same time, considerable variability is exhibited when decision-makers, judges and jurors alike, are asked to convert these qualitative, ordinal judgments into quantitative monetary awards (Saks et al. 1997; Diamond, Saks, and Landsman 1998; Kahneman, Schkade, and Sunstein 1998; Wissler, Hart, and Saks 1999). For instance, in a large-scale experiment involving more than 1,000 jury-eligible participants who viewed a videotape of a product liability trial, the standard deviation of the damages awarded was 138% of the mean for economic damages and 313% of the mean for pain and suffering damages. When analyzing trimmed values (where values above the 97th percentile were treated as if the jurors favored the award determined by jurors at the 97th percentile), the standard deviation was 75% for economic damages and 154% for pain and suffering damages (Diamond, Saks, and Landsman 1998).

The scope of this variability and whether it should be a cause of serious concern outside the laboratory are, however, debated. Commentators claim that damage awards,

for example, are largely predictable and sensible, taking into account the subtle differences in the characteristics of seemingly similar cases (Greene and Bornstein 2003; Vidmar and Hans 2007, pp. 299–302). Judges' power to review jury awards, appellate courts' supervision of trial judges' awards, and the fact that most claims are settled through negotiations managed by experienced attorneys—all significantly reduce the actual impact of the distorting factors observed in the laboratory (Eisenberg, Rachlinsky, and Wells 2002). Moreover, it may be argued that some apparently irrelevant anchors (see section 7.5 below), such as the effect of economic damages on punitive damages, are not normatively irrelevant (Eisenberg, Rachlinsky, and Wells 2002). And yet the overall picture emerging from experimental and empirical research, supported by the common sentiment expressed in the legal community, is that jury-set monetary awards, especially for noneconomic and punitive damages, are unjustifiably variable and irregular.

### 7.3 Individual and Group Decision-Making

While some judicial decisions are made by a single person, many are made by a panel of three or more judges, and all jury decisions are a product of group deliberation (see section 11 below). There is mixed evidence as to whether group deliberation increases the coherence and predictability of quantitative decisions. On the one hand, Diamond, Saks, and Landsman (1998) found that the standard deviation of total individual awards, prior to deliberation, was over \$7,000,000. Following deliberation in groups of six, the standard deviation of total jury awards dramatically dropped to less than \$1,000,000. The dramatic decrease in variability was exhibited in economic and non-economic damages alike.

On the other hand, in a very large experimental study involving more than 3,000 subjects, Schkade, Sunstein, and Kahneman (2000) found that at least in the context of punitive damages, jury deliberation actually reduced predictability. Due to the well-documented phenomenon of group polarization, jury dollar verdicts were systematically higher than median predeliberation judgments (see also Diamond and Casper 1992, pp. 553–57). Since this tendency was more pronounced when the median of jurors' predeliberation judgments was high, the overall variability of the awards increased. Among juries that awarded punitive damages, 27% awarded sums that were as high as, or higher than, the highest predeliberation judgment of their individual members (see also Sunstein et al. 2002). Further research is necessary for explaining the contradictions in these findings and delineating the opposite possible effects of jury deliberation.

### 7.4 Models of Quantitative Judicial Decision-Making

Several models have been proposed to describe the cognitive process of deriving numerical values from qualitative assessments. Kahneman, Schkade, and Sunstein

(1998) proposed a descriptive model of the process by which individual jurors set punitive damages, dubbed the “outrage model.” According to the model, outrage results from evaluation of the defendant’s behavior. Combined with the ensuing harm, outrage stimulates the intent to punish. Once intent to punish is formed, jurors express this attitude by transforming intent into a dollar scale. Since there is no obvious way to conduct such a transformation, the process is prone to strong influence by various anchors (see section 7.5 below). Hastie (2011) proposed a more general, four-stage model—the “intention + anchor model”—that applies to determination of other numerical verdicts as well.

These and related models (such as the one proposed by Hans and Reyna 2011) leave open the question of the order of damage determination. Do decision-makers first calculate damages for each category or subcategory of harms and losses separately, and then add up the numbers, or do they first determine a global award and then—if required to do so—break down the total into the different categories. The available empirical and experimental data appear to indicate that both mechanisms come into play. Decision-makers engage in a certain amount of calculation, yet their intuition about the appropriate total award is important as well (Greene and Bornstein 2003, pp. 159–61).

## 7.5 Anchoring

Both the experimental findings and the theories explaining them point to the centrality of anchors. When people are presented with a salient number before they make a numerical judgment, they tend to make their judgment through adjustments from the initial number, which serves as an anchor. The adjustment, however, is often insufficient and therefore the judgment is biased towards the anchor (Chapman and Johnson 2002). In serially considering factors that may require adjusting the initial number, there is a tendency to underuse this information and to halt the adjustment process too early (Hastie 2011; see also Epley and Gilovich 2006). The anchor also biases information sampling. It draws people’s attention to information that is consistent with the initial anchor and away from information that would lead to greater adjustment (Hastie 2011). Furthermore, the closer a factor is to the initial anchor, the more the decision-maker is likely to focus on the similarity between the two, and the further away a factor is from the initial anchor, the more the decision-maker is likely to focus on the dissimilarity, hence underweighting its relevance (Mussweiler 2003).

Studies have shown that numbers can serve as anchors even if they provide no meaningful information about the decision task, and even if decision-makers are fully aware of their meaninglessness (e.g., Tversky and Kahneman 1974; Englich, Mussweiler, and Strack 2006). When a problem has a right answer, being knowledgeable about the pertinent issue reduces or even eliminates the anchoring effect (Wilson et al. 1996). At the same time, anchoring influences both laypersons and experts, including judges (Guthrie, Rachlinski, and Wistrich 2001; Englich, Mussweiler, and Strack 2006).

Experiments have shown that it is difficult to overcome the anchoring effect even when forewarned, and that this bias is unlikely to be affected by monetary incentives to give the right answer when such an answer exists (Chapman and Johnson 2002; English, Mussweiler, and Strack 2006).

Experimental and empirical studies have highlighted the role of several anchors in the context of quantitative judicial decision-making. A common anchor is the *amount of economic damage*. Strong correlations have been found between economic and noneconomic damages and between compensatory and punitive damages (Eisenberg et al. 1997). These correlations may indicate that the former serve as an anchor when determining the latter (Sunstein et al. 2002; Hans and Reyna 2011). Since the severity of the harm is a relevant factor when determining economic, noneconomic, and punitive damages, this correlation may seem perfectly sensible; and even a direct inference from economic to noneconomic and punitive damages is not necessarily groundless (Sunstein et al. 2002; Eisenberg, Rachlinsky, and Wells 2002). At the same time, there is evidence that the amount of economic damages affects the amount of noneconomic damages more strongly among jurors than among judges (Hans and Reyna 2011). This finding arguably indicates that laypersons are overly influenced by the plaintiff's economic loss when determining noneconomic and punitive damages.

A more troubling anchor is the *amount of damages claimed by the plaintiff* (Chapman and Bornstein 1996; Hastie, Schkade, and Payne 1999; Viscusi 2001b; Greene and Bornstein 2003, pp. 151–55). Despite the fact that plaintiffs asking for extremely large amounts of compensation were perceived less favorably by the subjects in an experiment conducted by Chapman and Bornstein (1996), the amount requested served as an anchor affecting the damages awarded. The effect was linear even for extreme amounts. McAuliff and Bornstein (2010) found that the way the figures are presented to the jury also influences the award. An empirical study of actual trials and jury deliberations (Diamond et al. 2011) revealed a more nuanced picture, in which plaintiffs' very high claims, especially for nonmonetary harms, was sometimes perceived not only as irrelevant but also as outrageous and hence counterproductive. Some experimental studies have also noted the existence of a boomerang effect (Marti and Wissler 2000; Greene and Bornstein 2003, p. 153).

While it may be argued that the compensation requested is an indication of the scope of the harm suffered by the plaintiff, inasmuch as it serves as an anchor, its obvious manipulability is a cause of concern. This is especially so if the linear effect of the damages claimed manifests itself even when the decision-makers do not believe that the damages requested indicate the level of the plaintiff's suffering or her medical expenses, as demonstrated in Chapman and Bornstein's study (1996).<sup>4</sup>

<sup>4</sup> Even more surprisingly, the amount requested also affected the judgment of causality: the higher the amount, the higher the assessed probability that the defendant caused the plaintiff's injury.

Guthrie, Rachlinski, and Wistrich (2001) appear to have demonstrated the anchoring effect of another irrelevant factor: a meritless motion to dismiss a tort case because it did not meet the minimum threshold of damages required for jurisdiction. The meritless jurisdictional motion induced a significant decline in the damages awarded by the judges who were exposed to it.

A particularly intriguing anchoring effect has been observed in experiments that studied the impact of *caps* on damages. In one study, Hinsz and Indhal (1995) found that caps dramatically increased the median total award in a case referring to the death of two children. In another study, Saks and his colleagues (1997) examined the influence of caps on damages for pain and suffering. As regards a high-severity injury, when subjects were informed about the cap, it dramatically reduced both the mean and the variability of the awards. Regarding a medium-severity injury, the introduction of the cap slightly increased the mean and decreased the variability of the awards, but none of these effects was statistically significant. For a low-severity injury, however, the cap greatly increased both the mean and the variability of the awards. Caps are thus able to prevent mega awards for pain and suffering, but to the extent that they are meant to increase the predictability of noneconomic damages, they likely produce the opposite outcome due to their anchoring effect in cases of low-severity injuries.<sup>5</sup> Comparable results were obtained in an experimental study of caps on punitive damages (Robbennolt 1999). Not informing the jury about the existence of a cap (with the judge imposing it after the jury has set the damages) may ameliorate this concern (as well as the concern that the jury would evade the cap by increasing the damages for uncapped counts: Anderson and MacCoun 1999; Greene, Coon, and Bornstein 2001; Sharkey 2005). However, the likelihood that such caps will remain secret in the long run does not seem very high.

## 7.6 Concluding Remarks and Normative Implications

As this section has shown, when it comes to translating qualitative judgments into quantitative decisions, the decisions predictability decreases, their variability increases, and considerable differences between judges and juries emerge (Wissler, Hart, and Saks 1999). These differences most probably result from jurors' limited information regarding the customary awards and punishments as well as from the vagueness of the instructions they receive. In the absence of any reliable reference point, juries are forced to rely on questionable data, such as the amount of damages claimed by the plaintiff or the defendant's profits.

One way to cope with the special difficulties that jurors face in this respect is not to entrust such decisions to juries, but to judges, who are presumably familiar with customary awards and sentences. This route has been taken by most legal systems

<sup>5</sup> Moreover, inasmuch as there is a problem of undercompensation for high-severity injuries and overcompensation for low-severity injuries, caps exacerbate this problem (Saks et al. 1997).



around the world. Another way to circumvent the problem is to provide juries with clearer instructions, such as sentencing guidelines, the average and range of customary awards, and examples of verdicts handed down in similar cases (Chase 1995; Saks et al. 1997; Wissler, Hart, and Saks 1999). The lack of clear jury instructions regarding quantitative decisions in many US jurisdictions is rather puzzling, and the calls for reforms (e.g., Greene and Bornstein 2003, pp. 202–3) are quite convincing.

## 8 ATTITUDES TOWARDS PROBABILISTIC EVIDENCE

---

Behavioral scholars have demonstrated that decision-makers exhibit diverse attitudes towards probabilistic evidence. Whereas fact-finders are reluctant at times to base liability on such evidence, they quickly deduce erroneous conclusions in other instances. This section briefly sketches some of the main findings in this area (see also chapter 27 by Vars in this volume).

One of the early findings of cognitive psychology is that people tend to undervalue probabilistic information (Tversky and Kahneman 1974). In the legal context, researchers have shown that people can easily be maneuvered to undervalue the probative power of scientific evidence that ties a person to an act. Because within a population of one million people, ten thousand people share a trait that is attributed to 1% of the population, many tend to think that this trait has no legal relevance. It has thus been argued that people underuse associative evidence (Thompson and Schumann 1987).

Empirical research nonetheless suggests that attitudes towards statistical evidence are driven by more than mere misunderstanding of probabilities. In a seminal study, Gary Wells (1992) experimentally studied people's reluctance to assign liability based on naked statistical evidence. He hypothesized that "in order for evidence to have a significant impact on people's verdict preferences, one's hypothetical belief about the ultimate fact must affect one's belief about the evidence" (Wells 1992, p. 746). For instance, the fact that 80% of the buses in a certain town belong to the blue bus company and 20% to the gray bus company is insufficient to find the former liable for an accident caused by an unidentified bus, because the determination of liability would not change one's belief about the accuracy of the statistical data. The statistical data remains true whether or not a blue bus was involved in the accident. In contrast, when a weigh-station attendant testifies that according to his records, a blue bus was weighed in the nearby station just before the accident—thus tying that bus company to the accident—the determination of liability would more likely bear on the reliability of this testimony, even if the defendant had already established that those records were wrong 20% of the time. In this case, determining which bus was involved in the accident does bear on the accuracy of the weigh-station's records and the reliability of the attendant's

testimony. Wells found that fact-finders are much more likely to assign liability in the second setting. This preference did not stem from any difficulty in dealing with probabilities, as subjects in both conditions accurately assessed the probability that the accident was caused by a blue bus.

Wells's hypothesis was challenged by subsequent studies that offered competing hypotheses and explanations for the "Wells Effect." For example, building on Kahneman and Tversky's (1982b) simulation heuristic, Niedermeier, Kerr, and Messé (1999) argued that the willingness to ground liability on statistical evidence depends on how easily one can imagine an alternative scenario that would be compatible with the evidence. In their experiments, subjects read various vignettes describing a lawsuit relating to an accident similar to the one analyzed by Wells. Keeping the probability constant, they used vignettes that differed with respect to the ease in which one could imagine a counterfactual scenario in which the defendant was not responsible for the harm done. The results showed that while the assessed probability was similar under the two conditions, willingness to accept the claim declined considerably under conditions conducive to imagining an alternative scenario.

In a recent study, Zamir, Ritov, and Teichman (2014) identified a general disinclination to base liability on circumstantial evidence that goes beyond the statistical nature of some circumstantial evidence, what they called an "anti-inference bias." In their experiments, subjects analyzed situations in which the probability of wrongdoing was held constant, yet the type of evidence was randomized between direct and circumstantial. For instance, people were more willing to assign liability for a speeding violation when it was detected by a single speed camera than when it was detected by a system of two cameras, located at two points on a toll road, which documented the precise time that the driver drove between them but not the actual speed.

Thus far we have focused on the reluctance of judicial decision-makers to base liability on certain types of evidence. Yet, other aspects of human decision-making might cause people to assign liability even when it is unwarranted. Special attention in this regard has been given to the phenomenon of base-rate neglect.

Base-rate neglect refers to people's tendency to discount information about the frequency with which the respective event occurs and focus instead on available individuating information. The following famous example from Kahneman and Tversky (1973) helps explain the phenomenon: Jack is a 45-year-old man. He is married with four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies, which include home carpentry, sailing, and mathematical puzzles. Subjects in one condition in this study were told that Jack was randomly drawn from a pool of people composed of 70 engineers and 30 lawyers, while subjects in the other group were told that the pool was composed of 30 engineers and 70 lawyers. When asked to estimate what Jack does for a living, the results showed that notwithstanding the base rate, approximately 90% of participants in both groups assumed that Jack is likely to be an engineer.

In an early study documenting decisions across a wide array of fields, Bar-Hillel (1980) demonstrated that base-rate neglect can cause errors in the judicial context. The

subjects in her study were told that a hit-and-run accident involving a taxicab occurred at night, with 85% of the city's cabs being blue and the remaining 15% green. In court, an eyewitness testified that the cab involved in the accident was green. The court examined the witness's capabilities and reached the conclusion that he was correct 80% of the time (and wrong 20% of the time). Subjects were then asked to evaluate the probability that a green cab was in fact involved in the accident. The results showed that they focused solely on the witness's credibility rate: Their mode and median estimates of the probability that a green cab was the culprit were 80%. Calculating the actual probability, however, requires taking the underlying probability that the cab is green into account, and is thus only 41% (approximately 10% of the participants roughly approximated this answer).

In a later experimental study conducted with acting judges, Guthrie, Rachlinski, and Wistrich (2001) used a tort case in which a warehouse barrel inadvertently harmed a passerby. The legal question was whether the warehouse workers' negligence caused the accident, or whether another factor was involved. Participants in the study were informed that "(1) when barrels are negligently secured, there is a 90% chance that they will break loose; (2) when barrels are safely secured, they break loose only 1% of the time; (3) workers negligently secure barrels only 1 in 1,000 times" (Guthrie, Rachlinski, and Wistrich 2001, p. 808). Based on this information, participants were asked to estimate the probability of negligence, and were offered four probability ranges: 0%–25%; 26%–50%; 51%–75%; 76%–100%. While the precise answer is 8.3%, most participants did not choose the lowest option. That being said, 40% of participants did choose this option, a result that is better than those achieved by other populations in comparable studies (see Guthrie, Rachlinski, and Wistrich 2007, pp. 22–24).

## 9 PRIMING AND PREJUDICE

---

Priming refers to "the incidental activation of knowledge structures, such as trait concepts and stereotypes, by the current situational context," leading to a particular cognitive or affective response (Bargh, Chen, and Burrows 1996, p. 230). Priming, by definition, occurs in implicit memory, which is accessible only indirectly (Bargh and Chartrand 2000). Retention of prior experiences within a knowledge structure can be revealed by measuring variance in the performance of certain tasks after exposure to a prime (the stimulus) relative to performance in the absence of such exposure. For example, in a classic study, Bargh, Chen, and Burrows (1996) primed half of their subjects with stereotypical traits of elderly people (by asking them to construct sentences with words such as *old*, *gray*, *forgetful*, and *wrinkle*), while the other half of their subjects was not primed (they were asked to construct sentences with neutral words such as *clean* and *private*). The study's dependent variable was the time it took participants to walk down a hall once they had completed writing their first set of sentences. Interestingly, participants exposed to the stereotypical prime walked more slowly than

did participants who received the neutral treatment. For a general overview of the phenomenon, see Moskowitz (2005).

Priming can have various implications for judicial decision-making, ranging from lawyers' litigation tactics and their ability to prime jurors and judges (Stanchi 2010), to the influence of religious and national symbols that are incorporated into the interior design of courtrooms (Hassin et al. 2007; Mazar, Amir, and Ariely 2008). This section does not discuss these implications, but rather focuses on the use of priming as an effective experimental procedure to determine whether subjects possess implicit racial biases—an issue of the utmost theoretical and normative importance.

Researchers have demonstrated that racially charged primes (e.g., rap versus pop music) can cause people to judge the behavior of blacks as more hostile when compared to other groups (Rudman and Lee 2002). More generally, scholars have employed the Implicit Association Test (IAT) to examine attitudes towards marginalized groups such as blacks, Asians, and homosexuals.<sup>6</sup> These studies have demonstrated that people hold many implicit biases towards different social groups and that these biases often predict behavior better than do explicit biases (Rudman and Lee 2002).<sup>7</sup>

Implicit bias research has been slowly trickling into legal analysis, including judicial decision-making (for a collection of studies, see Levinson and Smith 2012). In the interest of brevity, we focus our discussion exclusively on the issue of race and the adverse effects of implicit bias on black litigants in the United States.

Eberhardt and her colleagues (2004) documented an implicit bidirectional association between blacks and criminality. When subliminally primed with black male faces, subjects were quicker to recognize blurred images of items associated with crime (e.g., guns). Perhaps more surprisingly, when subliminally primed with images of items associated with crime, participants were more attendant to black male faces. Thus, as the authors note, "Not only are Blacks thought of as criminal, but also crime is thought of as Black" (Eberhardt et al. 2004, p. 883). Related findings demonstrating an implicit association between black males and guns have been reported in numerous studies that examined people's tendency to shoot in a video simulation involving armed and unarmed whites and blacks (Correll et al. 2002; Plant and Peruche 2005). However, while all these findings suggest that people tend to draw a connection between race and crime, a connection that might well carry into the courtroom, they were not conducted in the concrete context of judicial decision-making.

Several other studies have examined more directly the role of implicit racial bias in judicial contexts. Graham and Lowery (2004) asked a sample of police officers and juvenile probation officers to analyze vignettes describing a crime-related scenario. Unbeknownst to the participants, half were subliminally primed with words

<sup>6</sup> The IAT documents peoples' implicit associations by measuring their response time in a computerized task. For a detailed description of the methodology and a meta-analysis, see Greenwald et al. (2009).

<sup>7</sup> The use of the IAT to measure implicit bias is not a matter of consensus. For a discussion see Arkes and Tetlock (2004); Banaji, Nosek, and Greenwald (2004).

associated with blacks while the other half were primed with words lacking a common theme. Immediately afterwards, the officers read two ambiguous criminal scenarios and were asked to rate the hypothetical offender on several traits (e.g., hostility and maturity) and to assess the culpability, expected recidivism, and deserved punishment of offenders whose race remained unspecified. Finally, the officers completed a general attitudes and beliefs questionnaire about race. The results suggest that an implicit bias rather than explicit attitudes channeled participants' decisions. Participants in the race-primed group viewed the offender more negatively and were willing to punish him more harshly.

Levinson, Cai, and Young (2010) have introduced a new IAT that measured the association between whites/blacks and guilty / not guilty judgments. They discovered an implicit association between black and guilty. Furthermore, this association was indicative of the way their subjects analyzed the evidence in ambiguous cases. More recently, Levinson, Smith, and Young (2014) presented several troubling findings in the context of the death penalty. They showed that when examining willingness to impose the death penalty during the screening of capital juries, the process stacks those juries with people exhibiting a relatively strong implicit racial bias. They also found a connection between people's IAT score and their willingness to impose the death penalty. That is, people who exhibited a greater implicit bias were more willing to convict a black defendant relative to a white defendant.

A more nuanced analysis of this point was offered by Rachlinski and his colleagues (2009). This study, which involved presiding judges, employed a two-stage design. In the first stage, judges preformed a standard IAT to determine their racial preferences. The results showed the existence of a white preference among white judges but no racial preference among black judges. In the second stage, judges were asked to evaluate three vignettes describing ambiguous criminal cases. Prior to the first two vignettes, where the culprits' race remained blurred, half of the judges were subliminally subjected to a racial prime. In the third vignette, the race of the defendant was overtly manipulated (African American or Caucasian). In contrast to Graham and Lowery (2004), Rachlinski and his colleagues did not identify a main effect associated with the racial prime. The evaluations of judges who were primed did not differ significantly from the evaluations of judges who were not primed. That said, the researchers did identify a marginally significant effect of judges' IAT scores on their sentencing decisions. Judges with a white preference on the IAT gave harsher sentences to defendants when primed with black-associated words rather than with neutral words. Judges with a black preference on the IAT, on the other hand, gave lower sentences when primed with black-associated words rather than with neutral words. With respect to the third vignette in which race was explicitly operationalized, the authors could not identify any effect when analyzing the group of judges as a whole. However, further analysis did reveal a three-way interaction between IAT scores, the judge's race, and the defendant's race. Specifically, IAT scores were unconnected to the outcomes reached by white judges, whereas black judges with a black preference tended to acquit more often. While it is difficult to generalize these results, one conclusion does seem to stand

out: the explicitness of race matters. As the authors note, “when judges are aware of a need to monitor their own responses for the influence of implicit racial biases, and are motivated to suppress that bias, they appear able to do so” (Rachlinski et al. 2009, p. 1221). Arguably, decisions in the courtroom more closely resemble the scenario depicted in the third vignette; hence, the extent to which implicit biases actually influence real-world decisions remains unclear.

Implicit bias is an emerging field in the judicial decision-making context, and much work will have to be done before we fully understand the phenomenon’s impact. Additional research should explore the precise way in which implicit bias operates (if at all) in actual courtrooms where people are more likely to attempt to overcome their predispositions. As Kang and his colleagues (2012) acknowledge, “because of the incredible difficulties in research design, we do not have studies that evaluate implicit bias in real criminal trials” (p. 1146). Moreover, we currently have very little information on the way in which racial bias functions in the domain of civil litigation. After thoroughly examining implicit bias in this context, Kang and his colleagues did “concede that [their] claims about implicit bias influencing jury decision-making in civil cases are somewhat speculative and not well quantified” (2012, p. 1168). Finally, identification of effective interventions capable of ameliorating the effects of the bias is probably the ultimate goal of this research project and should be addressed (on potential policies see Roberts 2012; Wilkins 2012).

## 10 RULES VERSUS STANDARDS: CERTAINTY AND PREDICTABILITY

---

Legal norms are conventionally classified into rules and standards. Rules typically condition legal outcomes on the existence of easily ascertainable, limited number of facts. Determination of legal capacity solely by reference to age, and punishment of drivers who exceed a certain speed limit by a fixed fine are paradigmatic rules. Standards, in contrast, embody substantive objectives and values, such as reasonableness, good faith, and unconscionability. Judgments based on standards require examination of the entire set of circumstances and their assessment in light of the values the standard embodies (Kennedy 1976).

For centuries, it has been recognized that the primary virtue of rules, as opposed to standards, is their ability to curtail people’s discretion and enhance the law’s certainty and predictability. This common wisdom has been challenged by the American legal realists and Critical Legal Studies scholars, who doubted that general legal norms, detailed as they might be, can dictate the judicial outcome in any particular case (e.g. Llewellyn 1940; Singer 1988). One reason for this incredulity was that even a system of detailed rules allows the court to choose which rule will apply to any set of facts. Paradoxically, the more elaborate the system of rules, the broader



is the judge's discretion when determining which rule to apply in any specific case (see also Schlag 1985).

It took more than fifty years before experimental and empirical findings were brought to bear on this debate. In experiments conducted with law students and recent law school graduates, Sheppard (2012) and Sheppard and Moshirnia (2013) demonstrated that a simple, bright-line rule requiring some action to be performed within a specific time limit constrains decision-makers' discretion more than a vague standard of "reasonable time" (see also Feldman and Harel 2008).<sup>8</sup>

While these studies compared judicial decisions according to a standard to decisions under a single, simple rule, the real choice is very often between vague standards and an elaborate system of rules consisting of various distinctions, provisos, and exceptions. Two experimental studies examined the predictability of judgments under either a set of detailed rules or a few general standards. Specifically, in a series of large-scale experiments conducted with advanced-years law students, Ellinghaus, Wright, and Karras (2005) compared different models of legal norms: detailed rules, slightly less detailed rules, and very general, vague standards. Subjects read a description of a legal dispute and made a decision according to one of the models of legal norms. The certainty and predictability of the legal norms were measured by the degree of consensus among the separate verdicts, that is, the broader the consensus, the more certain and predictable the legal norms.

The main conclusion invited by these experiments was that systems of elaborate legal rules do not yield more certain and predictable outcomes than do systems of vague standards. Moreover, while standards produced more predictable outcomes in easy cases, the application of rules to relatively easy cases did not increase predictability (Ellinghaus, Wright, and Karras 2005, pp. 38–41).

In another experiment, Ellinghaus and his coauthors asked responders to rate the fairness of the outcome and the extent to which the judgment took the important facts of the case into account. It turned out that in easy cases decided according to standards, a positive correlation was found between the rated fairness of the judgment and the extent to which the responders believed that the judgment considered all the pertinent circumstances. No such correlation was found regarding judgments based on detailed rules. It thus appears that standards are superior for drawing decision-makers' attention to the more important aspects of a case.

In a follow-up study, Wright and his colleagues (2012) examined how the scope of data considered by the judge affected the judgment's certainty and predictability. In addition to making a judicial decision, the subjects in this experiment were asked to assess the importance of fifteen factual circumstances that, according to the experimenters'

<sup>8</sup> Sheppard and Moshirnia (2013) also demonstrated that legal argumentation had a greater effect on decisions made according to a rule than according to a standard. When deciding according to a vague standard, decision-makers can opt for the outcome they ideologically prefer with or without legal arguments. Legal arguments assist decision-makers reach their desired outcome when this outcome is in tension with the straightforward meaning of the rule.



judgment, belonged to one of three categories in terms of their importance. The participants were asked to provide this assessment four times during the experiment: after reading the factual description, after reading the description of the applicable legal norms, after writing the arguments favoring each side's position, and after writing their reasoned judgment. In this experiment the predictability and certainty of the legal norms were likewise determined by the degree of consensus regarding the decision.

The participants generally shared the experimenters' judgment as to the relative importance of the different circumstances. As the subjects progressed along the decision process, they ranked the important facts as more important and the unimportant facts as less important. However, contrary to the subjects deciding according to standards, who attributed increasing importance to the circumstances in the middle category, the subjects deciding according to specific rules attributed decreasing importance to intermediate circumstances. This finding supports the contention that the greater predictability and certainty of decisions under a regime of standards is due to decision-makers' increased attention to a larger set of circumstances. Wright and his colleagues also constructed a connectionist model of mental representations of the data presented to their subjects—the type of model underlying the coherence-based theories of legal decision-makers discussed in section 2. They showed that the phenomenon of coherence shift neatly explains their findings.

Considerably more experimental work is required to determine the generality of these findings. Strong support for the results can be found in a large comparative, empirical study of the certainty and consistency of enforcement mechanisms, under either legal rules or standards, conducted in the context of nursing-home regulation (Braithwaite and Braithwaite 1995). The study found much greater consistency among assessments made by supervisors under a system of standards than those based on a very detailed, intricate set of rules. It turned out that a multiplicity of technical rules give supervisors greater discretion in employing them and in choosing how much emphasis to put on each one.

## 11 GROUP DECISION-MAKING

---

Thus far we have focused mainly on judicial decision-making at the individual level. However, many judicial decisions are reached by a group. In this section we examine how group dynamics influence the outcomes of the judicial process.

A large body of psychological research has been dedicated to the question of whether group decision-making can overcome the errors associated with individual decision-making. As these studies show, group dynamics can elicit a wide variety of outcomes (for a review see Kerr, MacCoun, and Kramer 1996). The move from individual to group decision-making sometimes mitigates deviations from expected utility theory, sometimes has no effect on such deviations, and sometimes intensifies them.

The diverse effects of group deliberation should be unsurprising given the variety of group characteristics, decision procedures, and decision objects.

The effect of group deliberation also depends on the nature of the pertinent heuristic. When the issue is computational and the correct answer is easily demonstrable (and group members aware of the error have sufficient incentives to correct other members' error), collective decision-making can help overcome individual biases. In contrast, when the issue is complex or involves a value judgment, as is often the case with judicial decision-making, the deliberation process can drive the group towards extreme outcomes that do not reflect members' pre-deliberation preferences—a phenomenon known as group polarization.

Group polarization occurs when an initial tendency of individual group members in one direction is strengthened following group discussion (Isenberg 1986; Myers and Lamm 1976; Sunstein 2000). The two primary explanations for this phenomenon are social comparison and informational influences. According to the former, people strive to perceive themselves and to be perceived by others favorably. When observing the general tendency within the group, they adopt a position that is in the same direction but somewhat more extreme. According to the latter explanation, when group members are initially inclined in one direction, the number and persuasiveness of the arguments articulated in that direction during deliberation are greater than in the opposite direction, thus strengthening the initial tendency.

Researchers have documented the ways in which group polarization affects judicial decision-making. As the existing studies show, the phenomenon can lead to opposing outcomes, depending on the specific context. With respect to a jury's decision to convict or acquit, MacCoun and Kerr (1988) conducted a meta-analysis of the existing studies that point towards what they term a bias towards leniency. When there is no clear majority within the jury, deliberation process becomes skewed towards acquittal. Based on two additional experiments, the authors suggest that this result may stem from the unique standard of proof applied to this decision-setting: beyond a reasonable doubt. To the extent that this standard reflects a well-entrenched social and legal norm, jury members advocating acquittal may have an asymmetric advantage during deliberations, which helps them swing the other members towards their view.

Yet group polarization not only mitigates legal outcomes. As described in subsection 7.3., in a large-scale experiment, Schkade, Sunstein, and Kahneman (2000) examined the effect of group deliberations on decisions relating to punitive damages by comparing juror's pre-deliberation and post-deliberation determinations. The jury's dollar verdicts were typically higher, or even far higher, than the median pre-deliberation judgment of the individual jurors.

Although the normative implications of this finding are ambiguous given that the setting did not generate a clear benchmark as to what the appropriate amount of punitive damages might be, the study's findings relating to the uncertainty generated by group discussion raises a clear concern. As explained in subsection 7.3., the deliberation process that brought the group to extremes also significantly increased the uncertainty of its outcomes. Based on this result, the authors conclude that "deliberation is a

significantly poorer way of aggregating opinions than is statistical pooling at least if the goal is to decrease the arbitrary unpredictability of awards” (Schkade, Sunstein, and Kahneman 2000, p. 1160).

As noted above, this behavioral phenomenon depends on the background norms governing the group. As Schkade and his colleagues acknowledge, their findings cannot be automatically generalized. For example, the defendant in all their studies was a corporation, and it is unclear whether similar attitudes would have been expressed with respect to individual defendants. Furthermore, different societies hold distinct views towards legal issues such as punishment (Mayhew and Kesteren 2002). Hence, whereas in some societies the lenient approach may have an advantage, in other societies the punitive argument may have the upper hand. In addition, one should be cautious with respect to these findings when norms change over time. For instance, in a more recent study, Devine and his colleagues (2004) observed a severity rather than leniency effect among jurors. As the authors noted, this finding “could reflect an attitudinal shift on the part of jurors since the 1970s,” when most of the groundbreaking work in this area was conducted (p. 2089). Apparently, the subtleties attached to group decision-making provide endless room for further research.

## 12 JUDGES VERSUS LAYPERSONS

---

A general question regarding cognitive biases pertains to the extent to which expertise diminishes the effect of those biases on decision-making. In legal contexts, the question should be rephrased as to whether professional judges make the same mistakes that people lacking legal training are likely to make. In responding to this issue, we here review some of the main findings on judges’ susceptibility to cognitive biases while focusing our discussion on controlled experimental studies that sought to isolate this issue.<sup>9</sup>

The psychological research on expertise is somewhat ambiguous. In general, judgments can reflect true expertise if they are reached within a decision-making environment that (1) is regular and predictable; and (2) offers people an opportunity to learn these regularities (Kahneman and Klein 2009). It is therefore important that decision-makers receive in a timely fashion feedback as to the quality of their choices. Empirical studies have shown that while some experts exhibit resilience to various biases (e.g., Korobkin and Guthrie 1997; Guthrie 2003), others do not (e.g., Anderson, Lowe, and Reckers 1993; Northcraft and Neale 1987).

The broadest body of work dealing with the cognitive aspects of professional judges’ decisions was presented by Chris Guthrie, Jeffrey Rachlinski, and Andrew Wistrich

<sup>9</sup> For field studies documenting the similarities and differences between jurors and judges in the area of punitive damages, see Eisenberg et al. (2006); Hersch and Viscusi (2004).

(GRW). The major contributions in this literature include Guthrie, Rachlinski, and Wistrich (2001, 2007, 2009), Rachlinski, Guthrie, and Wistrich (2007, 2011), and Wistrich, Guthrie, and Rachlinski (2005). Throughout this chapter, we have cited many of their findings. At this point we would like to briefly highlight the “big picture” emerging from their studies of judicial behavior. According to GRW, judges are “generally susceptible to the heuristics and biases that tend to induce intuitive and impressionistic judgments” (Guthrie, Rachlinski, and Wistrich 2009, p. 1521). Like most people, judges exhibit a tendency to base decisions on quick intuitions rather than on more complex deliberation. Judges’ results on the Cognitive Reflection Test mirrored those of other well-educated individuals (Guthrie, Rachlinski, and Wistrich 2009, pp. 1495–1500). In numerous studies conducted with different groups of judges, GRW demonstrated that anchoring, hindsight, framing, and other documented biases influence the way in which judges analyze legal vignettes (e.g., Guthrie, Rachlinski, and Wistrich 2001). These results were replicated with generalists as well as with judges who specialize in a specific area of law (Rachlinski, Guthrie, and Wistrich 2007; Guthrie, Rachlinski, and Wistrich 2009).

GRW did, however, document the aptitude of judges to overcome some of the pitfalls of human decision-making. For example, while judges usually found it difficult to ignore inadmissible evidence (Wistrich, Guthrie, and Rachlinski 2005), they did succeed in doing so at times (Wistrich, Guthrie, and Rachlinski 2005; Guthrie, Rachlinski, and Wistrich 2009). Specifically, a substantial body of work has demonstrated that judges can successfully deal with the challenge of ignoring inadmissible evidence in the context of the determination of probable cause. Judges in the United States must examine whether a probable cause for a search exists either in foresight (for search warrant purposes) or in hindsight, when the outcome of the search is already known (for evidentiary purposes). In a study involving 900 state and federal judges, GRW showed that their subjects made similar rulings in both contexts (Rachlinski, Guthrie, and Wistrich 2011; see also Wistrich, Guthrie, and Rachlinski 2005; Guthrie, Rachlinski, and Wistrich 2007).

Although GRW clearly show that judges are influenced by cognitive biases, this finding still leaves open the question of how judges fare on this front when compared to jurors. Several studies have attempted to examine this question directly by using the same survey instruments with judges and mock jurors. Hastie and Viscusi (1998), for example, compared the extent to which the hindsight bias influenced the decisions of both groups (see also Anderson et al. 1995; Viscusi 1999). They found that while mock jurors exhibited a clear hindsight bias, judges only exhibited trends towards hindsight (that were mostly statistically insignificant). In another study, Viscusi (2001a) compared judges and jurors along numerous dimensions of tort litigation and again found “fewer biases by judges in their treatment of risk” (p. 110). Judges were more open to conducting unbiased cost-benefit analyses of precautions and tended to perceive risk more accurately. In contrast, jurors more strongly exhibited a “zero-risk mentality” (Viscusi 2001a, p. 130), perhaps the product of a certainty effect (see chapter 13

by Williams in this volume), and were more willing to spend unlimited amounts of money to eliminate small risks.

### 13 A GENERAL ASSESSMENT OF BEHAVIORAL RESEARCH OF JUDICIAL DECISION-MAKING

---

The behavioral study of judicial decision-making faces unique challenges. Judges are, of course, human beings. But they are also trained jurists and professional adjudicators. A fundamental task of behavioral research is to examine how, if at all, legal training and judicial experience affect how judges make decisions. Making claims about judicial decision-making based on the findings of experiments conducted with laypersons is intrinsically problematic.

In this context, Schauer (2010) has claimed that a distinction should be drawn between tasks that both judges and other people (including lay jurors) perform, such as fact-finding and verdict-rendering, and tasks that lie within judges' exclusive province: selecting, interpreting, applying, and developing legal norms. Even if judges' decision-making is not fundamentally different from others' when performing nonexclusive tasks, there is special interest in examining how judges perform their uniquely designated tasks. Possibly, judges' legal training and experience, self-selection to become judges, and the institutional environment in which they operate, make a difference in this regard. In fact, so argues Schauer, if there is no significant difference between "thinking as a lawyer" or "reasoning as a judge," and thinking and reasoning as a layperson, then there is no reason to investigate judicial decision-making more than there is to investigate mechanics' or dentists' decision-making. Schauer further claims that current behavioral research of judicial decision-making leaves much to be desired.

While behavioral research of judicial decision-making is indeed in a relatively early stage of its development, Schauer's critique appears overstated for several reasons. First, as detailed in section 12, a considerable number of studies have used professional judges as subjects. There have also been some experimental studies of the "exclusive" judicial task of applying rules and standards to given sets of facts (see section 10). At the same time, we should concede that even laboratory experiments using professional judges as subjects differ strikingly from the real-world performance of judges; hence, any inference drawn from the former to the latter must be considered with caution (Vidmar 2011).

Inasmuch as judicial functions are fulfilled by juries, a huge body of empirical and experimental research has dwelt on jury decision-making, using jury-eligible people as subjects and observing actual jury deliberations (for an overview see Vidmar and Hans 2007; see also Greene and Bornstein 2003; Bornstein et al. 2008). Moreover, inasmuch as the hypothesized difference between judges and laypersons rests on the former's legal training, experiments conducted with advanced-year law students and experienced advocates provide additional, relevant insights (see, e.g., Ellinghaus, Wright, and Karras 2005; Zamir and Ritov 2012; Zamir, Ritov, and Teichman 2014).

A considerable number of studies have also directly compared judges and laypersons. As discussed in section 12, some of these studies found certain differences between professional judges and the general population (e.g., Wistrich, Guthrie, and Rachlinski 2005), while others found no divergence (e.g., Landsman and Rakos 1994). Even if no significant differences between judges' and other people's decision-making exist, there is much to be gained from examining how general psychological phenomena interact with the unique procedural and institutional characteristics of judicial decision-making (Sherwin 2010). Such an examination could lead to reforms in institutional design, court procedures, and even substantive legal rules. For instance, if judges are reluctant to impose liability based on circumstantial evidence, the law should perhaps redefine the constitutive elements of liability in a way that would obviate the need to make inferences from this type of evidence (Zamir, Ritov, and Teichman 2014). Just as the large corpus of behavioral research on physicians' decision-making informs the operation of health systems, there is a need for similar research in the judicial sphere, irrespective of whether or not judges differ from other decision-makers.

Finally, there is often additional support for the external validity of laboratory experiments of judicial decision-making. This includes studies conducted with experienced professionals in other domains, judges' self-reported descriptions of judging, and analyses of actual judgments (Simon 2010).

At the end of the day, one should concede that further research is necessary to establish the validity and generality of behavioral claims about judicial decision-making. This need is particularly conspicuous regarding judges' unique tasks of interpreting, developing, and applying legal norms. There is also room for qualitative research of judges' actual behavior in court as a method for ascertaining the external validity of some laboratory findings (Vidmar 2011). At the same time, much has already been achieved in this sphere. Furthermore, behaviorally informed theories and policy recommendations, resting on imperfect experimental and empirical data, are generally preferable to theories and recommendations resting on no such data.

## ACKNOWLEDGMENTS

---

We thank Ori Katz and Maya Zamir for excellent research assistance. This research was supported by the Israel Science Foundation (grants no. 100/11 and 1283/11) and the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation (grant No. 1821/12).

## REFERENCES

---

- Anderson, J. C., M. M. Jennings, S. E. Kaplan, and P. M. Reckers. 1995. The Effect of Using Diagnostic Decision Aids for Analytical Procedures on Judges' Liability Judgments. *Journal of Accounting and Public Policy* 14:33–62.



- Anderson, J. C., D. J. Lowe, and P. M. Reckers. 1993. Evaluation of Auditor Decisions: Hindsight Bias Effects and the Expectation Gap. *Journal of Economic Psychology* 14:711–37.
- Anderson, M. C., and R. J. MacCoun. 1999. Goal Conflict in Juror Assessments of Compensatory and Punitive Damages. *Law and Human Behavior* 23:313–30.
- Arkes, H. R., and C. Blumer. 1985. The Psychology of Sunk Costs. *Organizational Behavior and Human Decision Processes* 35:124–40.
- Arkes, H. R., and P. E. Tetlock. 2004. Attributions of Implicit Prejudice, or “Would Jesse Jackson ‘Fail’ the Implicit Association Test?” *Psychological Inquiry* 15:257–78.
- Bailey, M. A., and F. Maltzman. 2011. *The Constrained Court: Law, Politics, and the Decisions Justices Make*. Princeton, NJ: Princeton University Press.
- Banaji, M. R., B. A. Nosek, and A. G. Greenwald. 2004. No Place for Nostalgia in Science: A Response to Arkes and Tetlock. *Psychological Inquiry* 15:279–310.
- Bargh, J. A., and T. L. Chartrand. 2000. The Mind in the Middle: A Practical Guide to Priming and Automaticity Research. Pp. 253–85 in *Handbook of Research Methods in Social and Personality Psychology*, edited by H. T. Reis and C. M. Judd. New York: Cambridge University Press.
- Bargh, J. A., M. Chen, and L. Burrows. 1996. Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action. *Journal of Personality and Social Psychology* 71:230–44.
- Bar-Hillel, M. 1980. The Base-Rate Fallacy in Probability Judgments. *Acta Psychologica* 44:211–33.
- Baum, L. 2010. Motivation and Judicial Behavior: Expanding the Scope of Inquiry. Pp. 3–25 in *The Psychology of Judicial Decision Making*, edited by D. Klein and G. Mitchell. New York: Oxford University Press.
- Blanck, P. D., R. Rosenthal, A. J. Hart, and F. Bernieri. 1990. The Measure of the Judge: An Empirically-Based Framework for Exploring Trial Judges’ Behavior. *Iowa Law Review* 75:653–84.
- Bornstein, B. H., R. L. Wiener, R. Schopp, and S. L. Willborn (eds.). 2008. *Civil Juries and Civil Justice: Psychological and Legal Perspectives*. New York: Springer.
- Braithwaite, J., and V. Braithwaite. 1995. The Politics of Legalism: Rules versus Standards in Nursing-Home Regulation. *Social and Legal Studies* 4:307–41.
- Brewer, N., and G. L. Wells. 2011. Eyewitness Identification. *Current Directions in Psychological Science* 20:24–27.
- Chapman, G. B., and B. Bornstein. 1996. The More You Ask for the More You Get: Anchoring in Personal Injury Verdicts. *Applied Cognitive Psychology* 10:519–40.
- Chapman, G. B., and E. J. Johnson. 2002. Incorporating the Irrelevant: Anchors in Judgments of Belief and Value. Pp. 120–38 in *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by T. Gilovich, D. Griffin, and D. Kahneman. New York: Cambridge University Press.
- Chase, O. G. 1995. Helping Jurors Determine Pain and Suffering Awards. *Hofstra Law Review* 23:763–90.
- Choi, S. J., M. Gulati, and E. A. Posner. 2012. What Do Federal District Judges Want? An Analysis of Publications, Citations, and Reversals. *Journal of Law, Economics, and Organization* 28:518–49.
- Clayton, C. W., and H. Gillman (eds.). 1999. *Supreme Court Decision-Making: New Institutional Approaches*. Chicago: University of Chicago Press.



- Correll, J., B. Park, C. M. Judd, and B. Wittenbrink. 2002. The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals. *Journal of Personality and Social Psychology* 83:1314–29.
- Cox, M., and S. Tanford. 1989. Effects of Evidence and Instructions in Civil Trials: An Experimental Investigation of Rules of Admissibility. *Social Behaviour* 4:31–55.
- Cutler, B. L., and M. B. Kovera. 2011. Expert Psychological Testimony. *Current Directions in Psychological Science* 20:53–57.
- Daftary-Kapur, T., R. Dumas, and S. D. Penrod. 2010. Jury Decision-Making Biases and Methods to Counter Them. *Legal and Criminological Psychology* 15:133–54.
- Devine, D. J., K. M. Olafson, L. L. Jarvis, J. P. Bott, L. D. Clayton, and J. M. T. Wolfe. 2004. Explaining Jury Verdicts: Is Leniency Bias for Real? *Journal of Applied Social Psychology* 34:2069–98.
- Diamond, S. S., and J. D. Casper. 1992. Blindfolding the Jury to Verdict Consequences: Damages, Experts, and the Civil Jury. *Law & Society Review* 26:513–64.
- Diamond, S. S., M. R. Rose, B. Murphy, and J. Meixner. 2011. Damage Anchors on Real Juries. *Journal of Empirical Legal Studies* 8:148–78.
- Diamond, S. S., M. J. Saks, and S. Landsman. 1998. Juror Judgments about Liability and Damages: Sources of Variability and Ways to Increase Consistency. *DePaul Law Review* 48:301–26.
- Eberhardt, J. L., P. A. Goff, V. J. Purdie, and P. G. Davies. 2004. Seeing Black: Race, Crime, and Visual Processing. *Journal of Personality and Social Psychology* 87:876–93.
- Eisenberg, T., J. Goerdts, B. Ostrom, D. Rottman, and M. T. Wells. 1997. Predictability of Punitive Damages. *Journal of Legal Studies* 26:623–61.
- Eisenberg, T., P. L. Hannaford-Agor, M. Heise, N. LaFountain, G. T. Munsterman, B. Ostrom, and M. T. Wells. 2006. Juries, Judges, and Punitive Damages: Empirical Analyses Using the Civil Justice Survey of State Courts 1992, 1996, and 2001 Data. *Journal of Empirical Legal Studies* 3:263–95.
- Eisenberg, T., and V. P. Hans. 2009. Taking a Stand on Taking the Stand: The Effect of a Prior Criminal Record on the Decision to Testify and on Trial Outcomes. *Cornell Law Review* 94:1353–90.
- Eisenberg, T., J. J. Rachlinski, and M. T. Wells. 2002. Reconciling Experimental Incoherence with Real-World Coherence in Punitive Awards. *Stanford Law Review* 54:1239–71.
- Eisenberg, T., and M. Wells. 1993. Deadly Confusion: Juror Instructions in Capital Cases. *Cornell Law Review* 79:1–17.
- Ellinghaus, M. P., E. W. Wright, and M. Karras. 2005. *Models of Contract Law: An Empirical Evaluation of Their Utility*. Annandale, New South Wales: Themis Press.
- Englich, B., T. Mussweiler, and F. Strack. 2006. Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making. *Personality and Social Psychology Bulletin* 32:188–200.
- Epley, N., and T. Gilovich. 2006. The Anchoring and Adjustment Heuristic: Why the Adjustments Are Insufficient. *Psychological Science* 17:311–18.
- Epstein, L., and J. Knight. 1998. *Choices Justices Make*. Washington, DC: CQ Press.
- Epstein, L., W. M. Landes, and R. A. Posner. 2013. *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice*. Cambridge, MA: Harvard University Press.

- Fein, S., A. L. McCloskey, and T. M. Tomlinson. 1997. Can the Jury Disregard That Information? The Use of Suspicion to Reduce the Prejudicial Effects of Pretrial Publicity and Inadmissible Testimony. *Personality and Social Psychology Bulletin* 23:1215–26.
- Feldman, Y., and A. Harel. 2008. Social Norms, Self-Interest and Ambiguity of Legal Norms: An Experimental Analysis of the Rule vs. Standard Dilemma. *Review of Law and Economics* 4:81–126.
- Fischhoff, B. 1975. Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment under Uncertainty. *Journal of Experimental Psychology: Human Perception and Performance* 1:288–99.
- Frank, J. 1930. *Law and the Modern Mind*. New York: Brentano's.
- Gely, R. 1998. Of Sinking and Escalating: A (Somewhat) New Look at *Stare Decisis*. *University of Pittsburgh Law Review* 60:89–147.
- Graham, S., and B. S. Lowery. 2004. Priming Unconscious Racial Stereotypes about Adolescent Offenders. *Law and Human Behavior* 28:483–504.
- Greene, E., and B. H. Bornstein. 2003. *Determining Damages: The Psychology of Jury Awards*. Washington, DC: American Psychological Association.
- Greene, E., D. Coon, and B. H. Bornstein. 2001. The Effects of Limiting Punitive Damage Awards. *Law and Human Behavior* 25:217–34.
- Greene, E., and M. Dodge. 1995. The Influence of Prior Record Evidence on Juror Decision-Making. *Law and Human Behavior* 19:67–78.
- Greenwald, A. G., T. A. Poehlman, E. L. Uhlmann, and M. R. Banaji. 2009. Understanding and Using the Implicit Association Test: III. Meta-analysis of Predictive Validity. *Journal of Personality and Social Psychology* 97:17–41.
- Guilbault, R. L., F. B. Bryant, J. H. Brockway, and E. J. Posavac. 2004. A Meta-analysis of Research on Hindsight Bias. *Basic and Applied Social Psychology* 26:103–17.
- Guthrie, C. 2003. Panacea or Pandora's Box? The Costs of Options in Negotiation. *Iowa Law Review* 88:601–54.
- Guthrie, C., and T. E. George. 2005. The Futility of Appeal: Disciplinary Insights into the “Affirmance Effect” on the United States Courts of Appeals. *Florida State University Law Review* 32:357–85.
- Guthrie, C., J. J. Rachlinski, and A. J. Wistrich. 2001. Inside the Judicial Mind. *Cornell Law Review* 86:777–830.
- Guthrie, C., J. J. Rachlinski, and A. J. Wistrich. 2007. Blinking on the Bench: How Judges Decide Cases. *Cornell Law Review* 93:1–41.
- Guthrie, C., J. J. Rachlinski, and A. J. Wistrich. 2009. The “Hidden Judiciary”: An Empirical Examination of Executive Branch Justice. *Duke Law Journal* 8:1477–530.
- Guttel, E. 2004. Overcorrection. *Georgetown Law Journal* 93:241–83.
- Hans, V. P., and V. F. Reyna. 2011. To Dollars from Sense: Qualitative to Quantitative Translation in Jury Damage Awards. *Journal of Empirical Legal Studies* 8:120–47.
- Hassin, R. R., M. J. Ferguson, D. Shidlovski, and T. Gross. 2007. Subliminal Exposure to National Flags Affects Political Thought and Behavior. *Proceedings of the National Academy of Science, PNAS* 104:19757–61.
- Hastie, R. 1993. Introduction. Pp. 3–41 in *Inside the Juror: The Psychology of Juror Decision Making*, edited by R. Hastie. Cambridge: Cambridge University Press.
- Hastie, R. 2011. The Challenge to Produce Useful “Legal Numbers.” *Journal of Empirical Legal Studies* 8:6–20.

- Hastie, R., D. A. Schkade, and J. W. Payne. 1999. Juror Judgments in Civil Cases: Effects of Plaintiff's Requests and Plaintiff's Identity on Punitive Damage Awards. *Law and Human Behavior* 23:445-70.
- Hastie, R., and W. K. Viscusi. 1998. What Juries Can't Do Well: The Jury's Performance as a Risk Manager. *Arizona Law Review* 40:901-21.
- Hersch, J., and W. K. Viscusi. 2004. Punitive Damages: How Judges and Juries Perform. *Journal of Legal Studies* 33:1-36.
- Hinsz, V. B., and K. E. Indhal. 1995. Assimilation to Anchors for Damage Awards in a Mock Civil Trial. *Journal of Applied Social Psychology* 25:991-1026.
- Holyoak, K. J., and D. Simon. 1999. Bidirectional Reasoning in Decision Making by Constraint Satisfaction. *Journal of Experimental Psychology: General* 128:3-31.
- Isenberg, D. J. 1986. Group Polarization: A Critical Review and Meta-Analysis. *Journal of Personality and Social Psychology* 50:1141-51.
- Jois, G. U. 2009. Stare Decisis Is Cognitive Error. *Brooklyn Law Review* 75:63-141.
- Kahneman, D., and G. Klein. 2009. Conditions for Intuitive Expertise: A Failure to Disagree. *American Psychologist* 64:515-26.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1991. The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives* 5:193-206.
- Kahneman, D., D. Schkade, and C. R. Sunstein. 1998. Shared Outrage and Erratic Awards: The Psychology of Punitive Damages. *Journal of Risk and Uncertainty* 16:49-86.
- Kahneman, D., and A. Tversky. 1973. On the Psychology of Prediction. *Psychological Review* 80:237-51.
- Kahneman, D., and A. Tversky. 1982a. The Psychology of Preferences. *Scientific American* 246:16-73.
- Kahneman, D., and A. Tversky. 1982b. The Simulation Heuristic. Pp. 201-8 in *Judgment under Uncertainty: Heuristics and Biases*, edited by D. Kahneman, P. Slovic, and A. Tversky. Cambridge: Cambridge University Press.
- Kamin, K. A., and J. J. Rachlinski. 1995. Ex Post ≠ Ex Ante: Determining Liability in Hindsight. *Law and Human Behavior* 19:89-104.
- Kang, J., M. Bennett, D. Carbado, P. Casey, N. Dasgupta, D. Faigman, R. Godsil, A. Greenwald, J. Levinson, and J. Mnookin. 2012. Implicit Bias in the Courtroom. *UCLA Law Review* 59:1124-186.
- Kelman, M., Y. Rottenstreich, and A. Tversky. 1996. Context-Dependence in Legal Decision Making. *Journal of Legal Studies* 25:287-318.
- Kennedy, D. 1976. Form and Substance in Private Law Adjudication. *Harvard Law Review* 89:1685-778.
- Kerr, N. L., R. J. MacCoun, and G. P. Kramer. 1996. Bias in Judgment: Comparing Individuals and Groups. *Psychological Review* 103:687-719.
- Kerwin, J., and D. R. Shaffer. 1994. Mock Jurors versus Mock Juries: The Role of Deliberations in Reactions to Inadmissible Testimony. *Personality and Social Psychology Bulletin* 20:153-62.
- Korobkin, R., and C. Guthrie. 1997. Psychology, Economics, and Settlement: A New Look at the Role of the Lawyer. *Texas Law Review* 76:77-142.
- Landsman, S., and R. Rakos. 1994. A Preliminary Inquiry into the Effect of Potentially Biasing Information on Judges and Jurors in Civil Litigation. *Behavioral Sciences and the Law* 12:113-26.
- Lee, D. H., D. A. Krauss, and J. Lieberman. 2005. The Effects of judicial Admonitions on Hearsay Evidence. *International Journal of Law and Psychiatry* 28:589-603.

- Levinson, J. D., H. Cai, and D. Young. 2010. Guilty by Implicit Racial Bias: The Guilty / Not Guilty Implicit Association Test. *Ohio State Journal of Criminal Law* 8:187–208.
- Levinson, J. D., and R. J. Smith (eds.). 2012. *Implicit Racial Bias across the Law*. Cambridge: Cambridge University Press.
- Levinson, J. D., R. J. Smith, and D. Young. 2014. Devaluing Death: An Empirical Study of Implicit Racial Bias on Jury-Eligible Citizens in Six Death Penalty States. *New York University Law Review* 89:513–81.
- Lieberman, J. D., and B. D. Sales. 1997. What Social Science Teaches Us about the Jury Instruction Process. *Psychology, Public Policy, and Law* 3:589–644.
- Llewellyn, K. N. 1940. On Reading and Using the Newer Jurisprudence. *Columbia Law Review* 40:581–614.
- MacCoun, R. J., and N. L. Kerr. 1988. Asymmetric Influence in Mock Jury Deliberation: Jurors' Bias for Leniency. *Journal of Personality and Social Psychology* 54:21–33.
- Mandel, G. 2006. Patently Non-obvious: Empirical Demonstration that the Hindsight Bias Renders Patent Decisions Irrational. *Ohio State Law Journal* 67:1391–463.
- Marti, M. W., and R. Wissler. 2000. Be Careful What You Ask for: The Effect of Anchors on Personal Injury Damages Awards. *Journal of Experimental Psychology: Applied* 6:91–103.
- Martinek, W. L. 2010. Judges as Members of Small Groups. Pp. 73–85 in *The Psychology of Judicial Decision Making*, edited by D. Klein and G. Mitchell. New York: Oxford University Press.
- Mayhew, P., and J. van Kesteren. 2002. Cross-National Attitudes to Punishment. Pp. 63–92 in *Changing Attitudes to Punishment*, edited by J. V. Roberts and M. Hough. Cullompton: Willan Publishing.
- Mazar, N., O. Amir, and D. Ariely. 2008. The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research* 45:633–44.
- McAuliff, B. D., and B. H. Bornstein. 2010. All Anchors Are Not Created Equal: The Effects of Per Diem versus Lump Sum Requests on Pain and Suffering Awards. *Law and Human Behavior* 34:164–74.
- McKenzie, C. R. M., S. M. Lee, and K. K. Chen. 2002. When Negative Evidence Increases Confidence: Change in Belief after Hearing Two Sides of a Dispute. *Journal of Behavioral Decision Making* 15:1–18.
- Mitchell, T. L., R. M. Haw, J. E. Pfeifer, and C. A. Meissner. 2005. Racial Bias in Mock Juror Decision-Making: A Meta-analytic Review of Defendant Treatment. *Law and Human Behavior* 29:621–37.
- Moskowitz, G. B. 2005. *Social Cognition: Understanding Self and Others*. New York: Guilford Press.
- Mussweiler, T. 2003. Comparison Processes in Social Judgment: Mechanisms and Consequences. *Psychological Review* 110:472–89.
- Mustard, D. B. 2001. Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts. *Journal of Law and Economics* 44:285–314.
- Myers, D. J., and H. Lamm. 1976. The Group Polarization Phenomenon. *Psychological Bulletin* 83:602–27.
- Niedermeier, K. E., N. L. Kerr, and L. A. Messé. 1999. Jurors' Use of Naked Statistical Evidence: Exploring Bases and Implications of the Wells Effect. *Journal of Personality and Social Psychology* 76:533–42.

- Northcraft, G. B., and M. A. Neale. 1987. Experts, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions. *Organizational Behavior and Human Decision Processes* 39:84–97.
- Pennington, N., and R. Hastie. 1986. Evidence Evaluation in Complex Decision Making. *Journal of Personality and Social Psychology* 51:242–58.
- Pennington, N., and R. Hastie. 1988. Explanation-Based Decision Making: Effects of Memory Structure on Judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14:521–33.
- Pennington, N., and R. Hastie. 1991. A Cognitive Theory of Juror Decision Making: The Story Model. *Cardozo Law Review* 13:519–57.
- Pennington, N., and R. Hastie. 1992. Explaining the Evidence: Tests of the Story Model for Juror Decision Making. *Journal of Personality and Social Psychology* 62:189–206.
- Pickel, K. L. 1995. Inducing Jurors to Disregard Inadmissible Evidence: A Legal Explanation Does Not Help. *Law and Human Behavior* 19:407–24.
- Plant, E. A., and B. M. Peruche. 2005. The Consequences of Race for Police Officers' Responses to Criminal Suspects. *Psychological Science* 16:180–83.
- Posner, R. 1993. What Do Judges and Justices Maximize? (The Same Thing Everybody Else Does). *Supreme Court Economic Review* 3:1–41.
- Prentice, R. A., and J. J. Koehler. 2003. A Normality Bias in Legal Decision Making. *Cornell Law Review* 88:583–650.
- Rachlinski, J., S. Johnson, A. Wistrich, and C. Guthrie. 2009. Does Unconscious Racial Bias Affect Trial Judges? *Notre Dame Law Review* 84:1195–246.
- Rachlinski, J. J., and F. Jourden. 2003. The Cognitive Components of Punishment. *Cornell Law Review* 88:457–85.
- Rachlinski, J. J., C. Guthrie, and A. J. Wistrich. 2007. Heuristics and Biases in Bankruptcy Judges. *Journal of Institutional and Theoretical Economics* 163:167–86.
- Rachlinski, J. J., C. Guthrie, and A. J. Wistrich. 2011. Probable Cause, Probability, and Hindsight. *Journal of Empirical Legal Studies* 8:72–98.
- Rachlinski, J. J., A. J. Wistrich, and C. Guthrie. 2013. Altering Attention in Adjudication. *UCLA Law Review* 60:1586618.
- Read, S. J., and D. Simon. 2012. Parallel Constraint Satisfaction as a Mechanism for Cognitive Consistency. Pp. 66–86 in *Cognitive Consistency: A Fundamental Principle in Social Cognition*, edited by B. Gawronsky and F. Strack. New York: Guilford Press.
- Richards, M. J., and H. M. Kritzer. 2002. Jurisprudential Regimes in Supreme Court Decision-Making. *American Political Science Review* 96:305–20.
- Ritov, I., and J. Baron. 1990. Reluctance to Vaccinate: Omission Bias and Ambiguity. *Journal of Behavioral Decision Making* 3:263–77.
- Ritov, I., and J. Baron. 1992. Status-Quo and Omission Biases. *Journal of Risk and Uncertainty* 5:49–61.
- Robbennolt, J. K. 1999. Anchoring in the Courtroom: The Effects of Caps on Punitive Damages. *Law and Human Behavior* 23:353–73.
- Robbennolt, J. K., J. M. Darley, and R. J. MacCoun. 2003. Symbolism and Incommensurability in Civil Sanctioning: Decision Makers as Goal Managers. *Brooklyn Law Review* 68:1121–58.
- Roberts, A. 2012. (Re)forming the Jury: Detection and Disinfection of Implicit Juror Bias. *Connecticut Law Review* 44:827–82.
- Roese, N. J., and K. D. Vohs. 2012. Hindsight Bias. *Perspectives on Psychological Science* 7:411–26.



- Rudman, L. A., and M. R. Lee. 2002. Implicit and Explicit Consequences of Exposure to Violent and Misogynous Rap Music. *Group Processes and Intergroup Relations* 5:133–50.
- Saks, M. J., L. A. Hollinger, R. L. Wissler, D. L. Evans, and A. J. Hart. 1997. Reducing Variability in Civil Jury Awards. *Law and Human Behavior* 21:243–56.
- Samuelson, W., and R. Zeckhauser. 1988. Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty* 1:7–59.
- Schauer, F. 2010. Is There a Psychology of Judging? Pp. 103–20 in *The Psychology of Judicial Decision Making*, edited by D. Klein and G. Mitchell. New York: Oxford University Press.
- Schkade, D., C. R. Sunstein, and D. Kahneman. 2000. Deliberating about Dollars: The Severity Shift. *Columbia Law Review* 100:1139–75.
- Schlag, P. 1985. Rules and Standards. *UCLA Law Review* 33:379–430.
- Schuller, R. A. 1995. Expert Evidence and Hearsay: The Influence of “Secondhand” Information on Jurors’ Decisions. *Law and Human Behavior* 19:345–62.
- Schweitzer, M. 1994. Disentangling Status Quo and Omission Effects: An Experimental Analysis. *Organizational Behavior and Human Decision Processes* 58:457–76.
- Segal, J. A., and H. J. Spaeth. 1993. *The Supreme Court and the Attitudinal Model*. Cambridge: Cambridge University Press.
- Segal, J. A., and H. J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited*. Cambridge: Cambridge University Press.
- Sharkey, C. M. 2005. Unintended Consequences of Medical Malpractice Damages Caps. *New York University Law Review* 80:391–512.
- Sheppard, B. 2012. Judging under Pressure: A Behavioral Examination of the Relationship between Legal Decisionmaking and Time. *Florida State University Law Review* 39:931–1001.
- Sheppard, B., and A. Moshirnia. 2013. For the Sake of Argument: A Behavioral Analysis of Whether and How Legal Argument Matters in Decisionmaking. *Florida State University Law Review* 40:537–600.
- Sherwin, E. 2010. Features of Judicial Reasoning. Pp. 121–30 in *The Psychology of Judicial Decision Making*, edited by D. Klein and G. Mitchell. New York: Oxford University Press.
- Siegel, N. S. 1999. Sen and the Hart of Jurisprudence: A Critique of the Economic Analysis of Judicial Behavior. *California Law Review* 87:1581–608.
- Simon, D. 1998. A Psychological Model of Judicial Decision Making. *Rutgers Law Journal* 30:1–142.
- Simon, D. 2004. A Third View of the Black Box: Cognitive Coherence in Legal Decision Making. *University of Chicago Law Review* 71:511–86.
- Simon, D. 2010. In Praise of Pedantic Eclecticism: Pitfalls and Opportunities in the Psychology of Judges. Pp. 131–47 in *The Psychology of Judicial Decision Making*, edited by D. Klein and G. Mitchell. New York: Oxford University Press.
- Simon, D., L. B. Pham, Q. A. Le, and K. J. Holyoak. 2001. The Emergence of Coherence over the Course of Decision Making. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:1250–60.
- Simon, D., C. J. Snow, and S. J. Read. 2004. The Redux of Cognitive Consistency Theories: Evidence Judgments by Constraint Satisfaction. *Journal of Personality and Social Psychology* 86:814–37.
- Simonson, I., and A. Tversky. 1992. Choice in Context: Tradeoff Contrast and Extremeness Aversion. *Journal of Marketing Research* 29:281–95.
- Singer, J. W. 1988. Legal Realism Now. *California Law Review* 76:465–544.
- Spranca, M., E. Minsk, and J. Baron. 1991. Omission and Commission in Judgment and Choice. *Journal of Experimental Social Psychology* 27:76–105.

- Stanchi, K. 2010. The Power of Priming in Legal Advocacy: Using the Science of First Impressions to Persuade the Reader. *Oregon Law Review* 89:305–50.
- Staw, B. M., and J. Ross. 1989. Understanding Behavior in Escalation Situations. *Science* 246:216–20.
- Stebly, N., H. M. Hosch, S. E. Culhane, and A. McWethy. 2006. The Impact on Juror Verdicts of Judicial Instruction to Disregard Inadmissible Evidence: A Meta-Analysis. *Law and Human Behavior* 30:469–92.
- Sunstein, C. R. 2000. Deliberative Trouble? Why Groups go to the Extreme. *Yale Law Journal* 110:71–119.
- Sunstein, C. R., D. Schkade, D. Kahneman, and I. Ritov. 2002. Predictably Incoherent Judgments. *Stanford Law Review* 54:1153–215.
- Tanford, S., and M. Cox. 1988. The Effects of Impeachment Evidence and Limiting Instructions on Individual and Group Decision Making. *Law and Human Behavior*. 12:477–97.
- Thompson, W. C., and E. L. Schumann. 1987. Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy. *Law and Human Behavior* 11:167–87.
- Tversky, A., and D. Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185:1124–31.
- Vidmar, N. 2011. The Psychology of Trial Judges. *Current Directions in Psychological Science* February 20:58–62.
- Vidmar, N., and V. P. Hans. 2007. *American Juries: The Verdict*. Amherst, NY: Prometheus Books.
- Viscusi, W. K. 1999. How Do Judges Think about Risk? *American Law and Economics Review* 1:26–62.
- Viscusi, W. K. 2001a. Jurors, Judges, and the Mistreatment of Risk by the Courts. *Journal of Legal Studies* 30:107–42.
- Viscusi, W. K. 2001b. The Challenge of Punitive Damages Mathematics. *Journal of Legal Studies* 30: 313–50.
- Weinshall-Margel, K. 2011. Attitudinal and Neo-institutional Models of Supreme Court Decision Making: An Empirical and Comparative Perspective from Israel. *Journal of Empirical Legal Studies* 8:556–86.
- Wells, G. L. 1992. Naked Statistical Evidence of Liability: Is Subjective Probability Enough? *Journal of Personality and Social Psychology* 62:739–52.
- Wilkins, P. A. 2012. Confronting the Invisible Witness: The Use of Narrative to Neutralize Capital Jurors' Implicit Racial Biases. *West Virginia Law Review* 115:305–62.
- Wilson, T. D., C. E. Houston, K. M. Etling, and N. Breeke. 1996. A New Look at Anchoring Effects: Basic Anchoring and its Antecedents. *Journal of Experimental Psychology: General* 125:387–402.
- Wissler, R. L., A. J. Hart, and M. J. Saks. 1999. Decisionmaking about General Damages: A Comparison of Jurors, Judges, and Lawyers. *Michigan Law Review* 98:751–826.
- Wissler, R. L., and M. J. Saks. 1985. On the Inefficacy of Limiting Instructions: When Jurors Use Prior Conviction Evidence to Decide on Guilt. *Law and Human Behavior*. 9:37–48.
- Wistrich, A. J. 2012. The Evolving Temporality of Lawmaking. *Connecticut Law Review* 44:737–826.
- Wistrich, A. J., C. Guthrie, and J. J. Rachlinski. 2005. Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding. *University of Pennsylvania Law Review* 153:1251–345.



- Wright, E. W., M. P. Ellinghaus, N. M. Close, and A. Heathcote. 2012. The Effect of Rule Determinacy on Deciding Contract Disputes: Experimental Data and Network Simulation. Available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1884195](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1884195).
- Zamir, E., 2012. Loss Aversion and the Law. *Vanderbilt Law Review* 65:829–94.
- Zamir, E., and I. Ritov. 2012. Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation. *Journal of Legal Studies* 41:165–207.
- Zamir, E., I. Ritov, and D. Teichman. 2014. Seeing Is Believing: The Anti-inference Bias. *Indiana Law Review* 89:195–229.

## CHAPTER 27

---

# EVIDENCE LAW

---

FREDRICK E. VARS

### 1 INTRODUCTION

---

PSYCHOLOGY has long informed thinking on evidence law (Hutchins and Slesinger 1928; Smith 1942; Skolnick 1961; Woocher 1977; Mendez 1984; Tanford 1990; Wellborn 1991; Leo and Ofshe 1998; Robbennolt 2005). The first challenge of this chapter is therefore to narrow the scope. My focus will be the analysis of psychological phenomena that squarely call into question assumptions of normative economic and rational choice theory. This includes primarily, but not exclusively, heuristics and biases and related findings (Tversky and Kahneman 1974). Heuristics and biases are cognitive processes that expedite rationalization and analytic reasoning. These shortcuts reduce the costs of decision-making, but sometimes produce errors. Behavioral economics attempts to identify and measure how cognitive shortcuts lead to predictable variances from traditional rational choice theory.

Given this definition, the applicability of behavioral economics to the field of evidence law should be apparent. Evidence law primarily concerns what pieces of information should be allowed to be considered in the fact-finder's decision process. Of key importance is the truth-seeking objective of trial. Given that rationality is predictably compromised by heuristics and biases, the relation between relevant evidence and the heuristics and biases triggered by such evidence can effectuate a better understanding of and a better formulation of evidence law in general.

In particular, behavioral economics has been applied to evidence law in at least four different ways: (1) to explain or justify current law; (2) to argue that current law counteracts or fails to counteract the findings; (3) to advocate changes in practice under existing law; and (4) to argue for law reform. On the surface, the first two applications appear descriptive and the second two normative. Below the surface all four missions share common assumptions about the power and generalizability of the psychological findings. Even superficially explanatory assertions are premised on the normative view that psychology should inform our understanding of evidence law. Whether it

should, and how law and practice should be changed, are incredibly complex questions. Illustrating the complexity and suggesting ways forward are the goals of this chapter.

After the introduction, section 2 of this chapter discusses exemplars of each of the four descriptive and normative uses of behavioral findings outlined above. Several types of evidence and rules of evidence have received more sustained attention by behavioralists, cutting across the four types of arguments. Section 3 next takes up several of those types and rules of evidence: (1) eyewitness testimony; (2) experts; (3) standards of proof; and (4) subsequent remedial measures. The final and longest section, section 4, is devoted to character evidence, surveying the literature and offering a new illustration. The example is the exclusion of past offense evidence from sex offense trials, but the broader goal is to caution against simplistic application of psychology to law and to suggest directions for richer and more nuanced future research. Moving from description to prescription presents serious challenges.

## 2 DESCRIPTIVE AND NORMATIVE USES OF BEHAVIORAL INSIGHTS IN EVIDENCE LAW

---

### 2.1 Do Biases Explain or Justify Current Evidence Law?

Purely explanatory uses of behavioral findings are rare. One article argues that many guilty criminal defendants do not exercise their right to silence due to the “availability heuristic”: they underestimate risks irrationally perceived as remote (Seidmann and Stein 2000, p. 448). Being caught in a lie seems more remote than a certain appearance of guilt by remaining silent. Another example of the availability heuristic may be helpful: people read many news stories about particular families losing their homes to foreclosure. Because such instances are easy to recall, the newspaper readers overestimate the likelihood of foreclosure (Tversky and Kahneman 1973). The right-to-silence authors go on to make the normative claim that their game-theoretic model of the Fifth Amendment, including this feature, leads to more accurate determinations of guilt (Seidmann and Stein 2000, pp. 498–502).

The psychological phenomenon of overcorrection has been described as follows: “Subjects who are initially presented with a piece of information, but are later required to discount this information, generally tend to subtract too much” (Guttel 2004, p. 247). Slipping seamlessly from descriptive to normative, one commentator claims that avoiding overcorrection can help “explain” (Guttel 2004, p. 245) and “justify” (p. 261) “rules concerning discovery, the review powers of appellate courts, the exclusion of hearsay, and other evidentiary rules of the common law” (p. 245). The idea is to weed out weak evidence before its admission and later exclusion unduly influences

jurors. Some authors jump straight to justification: for example, cognitive biases such as availability and representativeness indicate that juries will overreact to evidence of the sexual conduct of victims of sexual assault, thereby justifying suppression of it in most cases (Kessler 1992, pp. 94–96).

Sometimes the claim is that current law reflects rather than combats biases in decision-making. Starting with the bias: “A person who follows [the representativeness] heuristic evaluates the probability of an uncertain event. . . by the degree to which it is (i) similar in essential properties to its parent population and (ii) reflects the salient features of the process by which it is generated” (Kahneman and Tversky 1982 p. 33). The classic example of where this strategy fails is overestimating the likelihood that a person described in terms similar to a stereotypical engineer is in fact an engineer even though he is drawn from a population containing relatively few engineers (Tversky and Kahneman 1982, p. 5). The representativeness heuristic has been linked to the legal doctrine of *res ipsa loquitur*: “According to the Restatement (Second) of Torts, when judging liability for an event that resulted in injury to a plaintiff, a fact-finder may infer that the defendant was negligent if ‘the event is of a kind which ordinarily does not occur in the absence of negligence’” (Rachlinski 2000, p. 90). Putting the two together, one author argues that, “[a]s it is ordinarily stated, *res ipsa loquitur* represents a profound misunderstanding of the laws of probability in precisely the way that the representativeness heuristic predicts” (Rachlinski 2000, p. 90; see also Kaye 1979). The obvious implication is that the doctrine should be changed.

Craig Callen (1994) offers one of the most sensitive, nonempirical applications of psychological findings to evidence law, involving the hearsay rule and its exceptions. Hearsay is a “statement, other than one made by the declarant while testifying at the trial or hearing, offered in evidence to prove the truth of the matter asserted” (Fed. R. Evid. 801). Callen shows that norms of cooperation in communication better explain existing doctrine than prevailing theories, but also justify some changes to that doctrine.

The lesson of these examples is that legal scholars generally cite behavioral findings in the service of normative claims. This no doubt reflects both the nature of legal scholarship generally and the fact that these findings are dubbed “biases,” “errors,” and “illusions.” Their very names would seem to justify correction.

## 2.2 Does the Adversarial Process Counteract Biases?

But perhaps the legal process already counteracts the effects of these biases. Some commentators think so, at least in certain areas (Froeb and Kobayashi 1996; Rachlinski 2000). Specifically, one commentator argues that the adversary process combats the tendency to give more weight to evidence that confirms a belief than to comparable evidence that refutes it (Bersoff 1992). Another commentator argues that the pretrial discovery of evidence and the settlement process reduce the impact of the overcorrection

bias (Guttel 2004). Others are less sure (Rachlinski 2011; Eisenberg, Rachlinski, and Wells 2001). A third group thinks that the legal system does not counteract biases in at least some areas—in particular, the strategic manipulation of fact-finders using psychology (Gold 1987) and the pervasive belief that what actually happened was inevitable and should have been foreseeable (i.e., the “hindsight bias”) (Hastie, Schkade, and Payne 1999). One commentator more radically asserts that cognitive limitations affirmatively *promote* the truth-seeking process by making it more difficult to lie successfully (Sanchirico 2004). Memory is more internally consistent, and thus resistant to cross-examination, than imagination. This diversity of opinion reflects the complexity of drawing normative, or even descriptive, conclusions for the legal system based on behavioral findings.

### **2.3 Should Practice under the Current Evidence Rules Be Changed?**

After demonstrating that judges are subject to a range of heuristics and biases, one group of commentators suggests that “judges might learn to educate themselves about cognitive illusions so that they can try to avoid the errors that these illusions tend to produce” (Guthrie, Rachlinski and Wistrich 2001). This modest proposal is followed by more aggressive ones, but it is important to note that simply learning about biases may help mitigate their effects (although the point is disputed). Not every behavioral finding justifies changes in the law.

### **2.4 Should the Rules of Evidence Be Changed?**

Because so few authors limit themselves to description (i.e., how a bias impacts fact-finding), this category overlaps substantially with the first category above (i.e., whether a bias justifies current evidence law). The only apparent difference is whether the commentator believes the status quo or a recommended alternative better counteracts a given bias. In addition to the other recommendations described in this chapter, one commentator has called for eliminating the exclusionary rule for evidence obtained from an illegal search and seizure and replacing it with a cause of action for damages (Slobogin 1999). Under current law, only individuals found to possess incriminating evidence seek suppression. This selection bias, combined with the representativeness and availability heuristics (described above), will lead judges to systematically deny suppression motions. Exposing judges to damage actions brought by innocent individuals would mitigate the biased inference of guilt (Slobogin 1999). Others suggest changes without expressly advocating them (Guthrie, Rachlinski, and Wistrich 2001; Jolls, Sunstein, and Thaler 1998).

## 3 CENTRAL APPLICATIONS OF BEHAVIORAL INSIGHTS TO EVIDENCE LAW

---

The following types and rules of evidence have received more sustained attention from commentators, implicating the same four descriptive and normative uses of behavioral evidence. Examining these uses one context at a time will hopefully solidify an understanding of these important pockets of evidence law and illustrate the evolution of thinking in each area.

### 3.1 Eyewitness Testimony

Fact-finders overvalue direct evidence, like eyewitness testimony, and undervalue circumstantial evidence (Heller 2006; see also Zamir, Ritov, and Teichman 2012). (Of course, this assumes that one knows the “correct” value of evidence, which usually will not be true. Controlled experiments can, however, provide good estimates in some domains.) The alleged culprits for erroneous weighting include the simulation (Heller 2006, pp. 259–64) and representativeness (Guthrie, Rachlinski, and Wistrich 2001) heuristics. One version of the simulation heuristic is estimating the probability of an event by trying to imagine a causal scenario; the easier it is to imagine such a scenario, the more probable the event appears (Heller 2006, p. 260). Some have prescribed changes in practice: more and better studies on witness identifications (Schacter et al. 2008) and reformed jury instructions (Heller 2006, pp. 304–5). Others more radically propose requiring corroboration for eyewitness testimony (Barzun 2008, p. 1993). Of these, the call for more research has the soundest footing given the limitations of present knowledge and complexity of translating findings into practice.

### 3.2 Experts

Expert testimony is perhaps the area of evidence law where the calls for change are most diverse. Commentators have argued that heuristics and biases justify (1) more expert testimony, (2) more careful screening of putative expert testimony by judges, and (3) replacing expert opinion testimony with data. This divergence in prescriptions reflects a peeling back of the onion: (1) lay jurors are biased; (2) judges are less biased than jurors; and (3) experts themselves, including judges, are subject to bias.

An early article in this line suggested that expert testimony combats lay jurors’ reliance on the representativeness heuristic and related neglect of base rates (recall the engineer example) (Saks and Kidd 1980, p. 134). It cited others’ calls for additional base-rate experts, special masters, court advisors, and special courts (p. 134). The

authors' own prescription was more modest: "experts ought to be permitted to offer their data, their algorithms, and their Bayesian [probability] theorems. The errors that may be introduced will be subject to adversarial cross-examination" (p. 148).

Notably, this article preceded the US Supreme Court's adoption in *Daubert v. Merrell Dow Pharmaceuticals* (1993) of a more active judicial gatekeeping role for expert testimony (Fed. R. Evid. 702). Behavioralists have largely embraced that role on the theory that judicial expertise and accountability can mitigate the ill effects of biases (Beecher-Monas 2003, p. 989; see also Gold 1987, pp. 510–12). This might appear like a flip from "more experts" to "fewer experts," but the new position can be described as an evolution toward "more and better experts" (for related suggestions to improve presentation of statistical evidence, see Lyon and Koehler 1996).

Perhaps not surprisingly, there have been new calls for additional experts, this time to help judges determine which expert testimony is reliable enough to be admitted. The rationale is that while judicial expertise is good, it is not good enough to evaluate scientific or other technical expert methodology. Such experts could avoid hindsight bias in particular, one set of commentators argue (Worthington et al. 2002).

Good experts are the solution to all problems, one might conclude from the foregoing. But experts are still human and thus prey to systematic bias (National Academy of Sciences 2009). Overconfidence and overoptimism are two particularly troubling tendencies for adjudication. Data may have other problems, but not these. One empirical study concludes: "In many settings, the fallible opinions of isolated experts should be supplemented or replaced by statistical data" (Meadow and Sunstein 2001, p. 631). The authors support this conclusion with data showing that experts substantially and systematically underestimate time to treatment for a particular medical condition (pp. 637–38).

The evolution of recommendations with respect to expert testimony illustrates a critical and sometimes neglected step in the behavioral economics of evidence law. The easy part is identifying a bias that affects laypeople. The hard part is crafting a solution that fixes more problems than it creates. An obvious first response to juror bias is to throw more sophisticated, and hopefully less biased, experts at the problem. But some of what they offer will be junk science and their opinions may be subject to the same or different biases that affect laypeople. Disqualifying bad experts (perhaps with the help of independent experts), or even replacing expert opinion with unbiased statistics, may end up being the best solution.

### 3.3 Standards of Proof

The standard of proof instructs fact-finders how to decide cases given uncertainty. Rational choice theory has an elegant solution to the question of where to set the standard of proof: find for the plaintiff if and only if the probability that he or she meets the elements of his or her case is above an algebraic combination of the utilities of the four possible outcomes (Kaplan 1968; Cullison 1969). The



details of the formula are not important for present purposes, but the four outcomes are, of course, true positive (i.e., a correct finding for the plaintiff), false positive, true negative, and false negative. Setting aside other criticisms (Tribe 1971; Guttel and Teichman 2012), one commentator argues from cognitive psychology results that standards of proof should be selected from a menu of no more than three, not from the continuum implied by rational choice theory (Clermont 1987, p. 1149). Selecting among standards is more complicated than the equation above (Vars 2010).

The two most studied existing standards are the criminal standard and the normal civil standard. In the common-law world, a criminal defendant should be convicted only if guilt is established “beyond a reasonable doubt.” In contrast, the plaintiff in a typical civil case should prevail if they make their case by a “preponderance of the evidence.” The latter standard has been equated to a probability greater than 50%, which is generally believed to minimize the total number of trial errors. However, behavioral research suggests that the “omission bias” leads individuals to demand a significantly higher probability (66%–75%) (Zamir and Ritov 2012, pp. 186–87). The omission bias is a tendency to favor inaction. In civil litigation, the researchers argue, accepting a claim is viewed as action, whereas rejecting a claim is inaction in that it preserves the legal status quo. The experimental finding of a higher threshold is important, but its normative implications are unclear (p. 197).

Along similar lines, studies have found that mock jurors’ verdicts are not impacted by verbal statements of the standard of proof (Kagehiro 1990). Quantifying the standards as probability thresholds, on the other hand, generated smaller win rates as the proof standard was raised. This might appear to be a victory for traditional rational choice theory, but at least one study has found verbal formulations effective and offered a psychological explanation. Rather than waiting until the end of trial to combine all of the evidence in Bayesian fashion, jurors begin trying to make sense of each piece of evidence as it is presented. Information supportive of the initial conclusion is then overvalued and conflicting information undervalued in what has been dubbed “a coherence shift” (Simon 2004, pp. 511–86; Engel 2009, p. 456). The tendency to overlook evidence of acquittal was apparently mitigated in an experimental setting by an actual beyond-a-reasonable-doubt instruction emphasizing subjective emotion, not probability: “you must be convinced of [his] [her] guilt to the same degree you would be convinced about a matter of importance in your own life in which you would act with confidence and without restraint or hesitation” (Engel 2009, p. 441, quoting Pennsylvania jury instruction). The study concludes that both the objectivists and subjectivists can learn from each other. The important point is that a behavioral understanding of juror decision-making may be essential in crafting standards of proof that accomplish their legal and social objectives (Korobkin and Ulen 2000, p. 1097; Jolls, Sunstein, and Thaler 1998, pp. 1429–30).

### 3.4 Subsequent Remedial Measures

People overestimate the predictability of past events. This tendency is known as “hindsight bias” (Fischhoff 1975). One set of commentators argue that excluding evidence of remedial measures taken after an injury to prove negligence before the injury (Federal Rules of Evidence 407) “represents an adaptation to the effects of the hindsight bias” (Guthrie, Rachlinski and Wistrich 2001). It appears that the commentators are making both descriptive and normative claims: this is how the law came to be, and it is a good thing. Logic provides some support for both claims. The hindsight bias is a robust psychological phenomenon and would predictably lead fact-finders sometimes to conclude that a defendant should have predicted an injury because the defendant took measures after it to prevent its recurrence. The defendant should have seen it coming sooner, the biased fact-finder might reason. To be sure, updating probability estimates based on new data is perfectly rational, if done correctly (Sanchirico 2003, pp. 1197–200).

So far so good, but there’s a missing step, as one of the authors previously observed: “suppressing subsequent remedial measures in accident cases slightly reduces the pool of relevant evidence that the fact-finder has available to decide the case. If the effect of the hindsight bias is small, these adaptations might be overreactions to the problem of judging liability in hindsight” (Rachlinski 2000, p. 73). Dan Kahan picked up on this important observation, concluding that excluding or admitting subsequent remedial measures on a case-specific basis may be optimal (Kahan 2010).

Kahan’s argument is significant on its own terms, but also as an exemplar of careful application of psychological findings to law. The critical insight is that evidence of subsequent remedial measures makes negligence somewhat more likely, even if not by as much as the hindsight bias leads fact-finders to believe, so a per se rule of exclusion will generate erroneous outcomes. Kahan, while expressly eschewing explanation, offers the following prescription: “If a judge excludes [subsequent remedial measures] proofs in cases in which the plaintiff’s case is otherwise weak, and is receptive to admission of such proofs in cases in which the issue of the defendant’s substandard conduct is otherwise close, she will minimize the sum total of erroneous outcomes—either mistaken findings of liability or mistaken findings of nonliability—relative to a rule that categorically excludes or admits such proofs” (2010, p. 1637).

But Kahan cannot resist the gravitational pull of current practice when he argues that this is what judges do all the time in balancing probative value and prejudicial effect under Federal Rule of Evidence 403, which basically requires that the value of evidence be net neutral or positive to be admissible (pp. 1638–39). Here Kahan falls short. He effectively assumes that the subsequent remedial measures evidence is the last piece of evidence offered so that “the full evidentiary context” is otherwise complete (p. 1639). This may not always be true. Relatedly, Kahan also assumes that judges measure probative value in

terms of how likely evidence is to change the outcome (“contribution. . . to accurate fact-finding” [p. 1649]). More plausibly, judges ask how far new evidence moves the meter. If hindsight bias is strong enough and probative value fixed, then unfair prejudice may substantially outweigh probative value in every case: weak, close, or overwhelming. Kahan may or may not be correct in all of his conclusions, but at least he frames the question well.

## 4. CHARACTER EVIDENCE

---

In the behavioral economics literature, perhaps the most discussed rule of evidence is the exclusion of character evidence. For example, under New York law, “The prosecutor in a criminal case may not introduce, during her case in chief, evidence of the defendant’s general character or propensity for criminal behavior. . . [or] specific incidents of defendant’s bad conduct” (Barker and Alexander 2012). In many jurisdictions, numerous exceptions may in practice overwhelm the rule, but for simplicity this illustration concerns only the general rule.

Past behavior is the best predictor of future behavior, so this rule on its face would seem accuracy-defeating (Tillers 1998, p. 792). Scholars have offered numerous justifications for the exclusion of character evidence (Sanchirico 2003, pp. 1239–59). Only one type of justification is relevant here: for psychological reasons juries will systematically overweight character evidence. The fundamental attribution error and representativeness heuristic are the most commonly cited and persuasive psychological culprits (Sanchirico 2003, pp. 1242–43; see also Saks and Kidd 1980, p. 136; Gold 1983, pp. 525–30; Park 1996; Korobkin and Ulen 2000, pp. 1086–87). Representativeness has been discussed above. The fundamental attribution error is the tendency to overvalue personality-based explanations for behavior and undervalue the importance of the situation. But does the undue weight overwhelm the probative or other positive values?

To assess this question a framework is needed. Accuracy in trial outcomes is an important purpose of evidence law. Minimizing the total number of errors is perhaps a sufficient definition of accuracy in the civil context, but avoiding false convictions (FP) is generally considered more important than avoiding false acquittals (FN). This justifies the requirement that guilt be proved beyond a reasonable doubt, which in one large survey judges most often defined as 90% certainty (McCauliff 1982, p. 1325, table 2; but see Simon and Mahan 1971).

Notwithstanding this high standard of proof, sex crime conviction rates for people brought to trial in two recent years (2009 and 2011) in New York City were 86% and 89% (Doll 2012). As alluded to above, evidence of past offenses is frequently admitted despite the general bar on character evidence of this kind. But assume for illustration that it isn’t. What would be the impact of adding prior offense evidence (as Federal Rule of Evidence 413 now allows)? The probability of conviction for defendants with prior sex offenses admitted would be given by the following formula, a straightforward application of Bayes’s Rule:

$$P(\text{Conviction}|\text{Prior}) = \frac{P(\text{Prior}|\text{Conviction}) \times P(\text{Conviction})}{P(\text{Prior})}.$$

One study found that roughly 12% of sex offenders in prison had a prior sex offense, which serves as a decent proxy for the probability of a prior given conviction ( $P(\text{Prior}|\text{Conviction})$ ) (Greenfeld 1997, p. 22, fig. 23). Assume next that a particular sex offense defendant has the same likelihood of conviction as the group of sex offenders who went to trial in New York City in 2011 ( $P(\text{Conviction}) = 89\%$ ). The third term,  $P(\text{Prior})$ , is the toughest one to estimate. Sex offenders make up a tiny fraction of the population (p. 2), but the pool of individuals on trial for a sex offense is obviously not a random cross-section of the population. It is helpful to divide the pool by conviction or acquittal.

$$P(\text{Prior}) = P(\text{Conviction}) \times P(\text{Prior}|\text{Conviction}) + P(\text{Acquittal}) \times P(\text{Prior}|\text{Acquittal}).$$

The first two terms have already been defined at 89% and 12%, respectively. The probability of acquittal is 11% (1 minus 89%). We have successfully pushed back the problem, but we are still left with an unknown: the probability that a defendant who will be acquitted has a prior sex offense ( $P(\text{Prior}|\text{Acquittal})$ ). Given that the recidivism rate for sex offenders is orders of magnitude higher than the first offense rate (Langan, Schmitt, and Durose 2003, pp. 1, 2), it is almost certain that the unknown value is less than 12%. But by how much? Three possibilities will frame the issue.

First, assume that acquitted individuals are just as likely as convicts to have a prior offense, that is, 12%. Introducing prior sex offense evidence would have no impact on the probability of conviction. The evidence has no probative value. At the other extreme, assume that acquitted individuals have the same very small likelihood of a prior sex offense conviction as the general population. Admitting evidence of a prior sex offense in this second scenario would all but ensure conviction ( $p > 99\%$ ). Assuming a history of priors halfway between these extremes (6%) and evidence thereof would push the probability of conviction from 89% to around 94%. (Notably, the effect is not symmetric. In this third scenario, for example, introducing evidence of *no* priors would reduce the probability of conviction by less than a percentage point.)

Suppose this third, midpoint, scenario accurately reflects reality. If juries were good Bayesians and correctly gauged the probative value of past offenses—the latter of which here is to say, assigned the correct values to  $P(\text{Prior}|\text{Conviction})$  and  $P(\text{Prior}|\text{Acquittal})$ —then past offense evidence would increase correct convictions for past offenders and slightly increase correct acquittals for past nonoffenders. The net result would be an essentially costless improvement in trial accuracy. The problem is that juries are not good Bayesians in general and systematically overweight prior sex offense information in particular.

As alluded to above, people tend to neglect base rates (like the percentage of engineers in the population)—or, more precisely, to give base rates less weight than Bayes's Rule dictates. In the present context, that would mean assigning too little weight to the past offense evidence. The probative value of the past offense evidence turns on two

base rates:  $P(\text{Prior}|\text{Conviction})$  and  $P(\text{Prior}|\text{Acquittal})$ . In contrast, the other variable in the equation,  $P(\text{Conviction})$ , is based on event-specific evidence like eyewitness identification, forensics, and victim testimony. In the extreme case juries would give no weight to past offense evidence (scenario 1 above) and its exclusion or admission would make no difference in outcomes.

A more likely and more serious error in this context is overweight. As noted above, the fundamental attribution error and the availability and representativeness heuristics point in that direction. In fact, people vastly overestimate recidivism among sex offenders: according to one survey, the public believes 74% of sex offenders will commit another sex offense (Levenson et al. 2007, p. 13, table 2); in contrast, only 5.3% of released sex offenders were rearrested within three years for a sex crime (Langan, Schmitt, and Durose 2003, pp. 1, 2). Of course, that means people believe sex offenders are much more likely than nonoffenders to have a prior offense (assuming that convictions are closely correlated with guilt, this implies that  $P(\text{Prior}|\text{Conviction}) \gg P(\text{Prior}|\text{Acquittal})$ ). Again taken to the extreme, this would ensure conviction at trial of almost every sex offender with a prior offense where the other evidence justified only an 89% conviction rate (scenario 2 above).

This is the danger that leads most commentators to defend the exclusion of prior offense evidence (Saks and Kidd 1980; Korobkin and Ulen 2000, Gold 1983). If overweight is really this extreme, then the commentators are probably right. In the present illustration, admitting priors leads to over 99% convictions when the correct conviction rate is 94%. However, by not allowing priors, the actual conviction rate remains a suboptimal 89%. Adding 5% correct convictions and 5% incorrect convictions would be justified only if false convictions and false acquittals were weighted equally (Friedman 2003, p. 969). They are not, as the beyond-a-reasonable doubt standard implies. If, however, the degree of overweight were slight, the trade-off might make sense. The important point is that the mere fact of overweight does not decide the question (Friedman 2003; Sanchirico 2001, p. 1246; Kahan 2010, p. 1634), even in this very stylized example.

In the real world, a whole host of other considerations is relevant in deciding whether to admit character evidence. The rules of evidence serve goals beyond accuracy at trial and operate as part of a complex system. To take two examples: (1) if past offense evidence were admitted in every case, prosecutors would more aggressively prosecute alleged repeat offenders, but, (2) due to case selection, this may or may not mean more of them would go to trial. Presumably, more individuals with past offenses would be charged but more would accept plea bargains. The other evidence of guilt for individuals with clean records would likely need to be stronger to justify indictment and prosecution. All of these responses to a change in the rules would affect the variables in the illustration above, with uncertain impact on net social welfare.

The rules of evidence also affect primary behavior (Kaplow 2012). Increasing the likelihood of sanction for individuals with prior offenses would presumably deter some from committing additional sex offenses. How strong is this effect likely to be? And how many sex offenses must be prevented to justify one false conviction? Christopher

Sanchirico argues that allowing character evidence would have the opposite effect and would actually *increase* crime. This conclusion is apparently premised on the notion that evidence of a prior offense, standing alone, would be sufficient for a conviction (Sanchirico 2001, p. 1266). Theory alone cannot answer the question.

The important lesson of this example is that identifying a bias is merely the first step. The second step is estimating the magnitude of the effect. To determine whether the bias justifies current law or a change thereof requires consideration of how that effect impacts a complex fact-finding process. How many outcomes change? How do we value each type of outcome? How would a change in practice or the law affect primary behavior? The values question is normative, but the other questions are, or should be, empirical. Progress in the behavioral analysis of evidence law will be achieved through careful natural and controlled experimentation.

## 5 CONCLUSION

---

Other scholars have concluded that empiricism is the way forward for behavioral economics and law (Goldman 2003, p. 224). Sensitivity to context is critical. As one commentator colorfully urges, “empirical legal research should fashion itself after cartography, seeking to create detailed maps of legal behavior in context” (Mitchell 2003, p. 1147). This is perhaps nowhere truer than in the law of evidence. First, one cannot be sure without testing that a particular bias affects decision-makers in a particular situation. Second, any assessment of impact or proposed reform must recognize that the introduction of a piece of evidence is just one part of a complex process. “[B]ecause the rules of evidence and procedure are merely parts of the larger legal system, the collateral consequences of tinkering with some parts while leaving others untouched must be taken into account” (Wistrich, Guthrie, and Rachlinski 2005, p. 1330).

Even with careful methodology, success is by no means assured. When psychological findings conflict with fundamental principles of the legal system, the system may reject the findings, no matter how robust (Rachlinski 2011, pp. 1690–96). A final word of caution about legal empiricism, which is not unique to behavioral economics or to evidence law: in the legal setting, measuring outcomes is difficult or sometimes even impossible. A primary goal of the law of evidence is accurate decision-making, but how does one measure accuracy? One commentator claims that “one reason that truth is problematic as a legal goal is that the fit between the truth and a jury verdict is impossible to measure” (Moreno 2003, p. 1186). This may be true in some cases, but DNA exonerations and related research (Risinger 2007) demonstrate that assessing the accuracy of trial outcomes is at times attainable.



## REFERENCES

- Barker, Robert A. and Vincent C. Alexander (2012). *N.Y. Practice Series, Evidence in New York State and Federal Courts* 5, § 4:13.
- Barzun, Charles L. (2008). "Rules of Weight." *Notre Dame Law Review* 83(5): 1957–2017.
- Beecher-Monas, Erica. "Heuristics, Biases, and the Importance of Gatekeeping." *Michigan State Law Review* 2003(4): 987–1022.
- Bersoff, Donald N. (1992). "Judicial Deference to Nonlegal Decisionmakers: Imposing Simplistic Solutions on Problems of Cognitive Complexity in Mental Disability Law." *SMU Law Review* 46(2): 329–71.
- Callen, Craig R. (1994). "Hearsay and Informal Reasoning." *Vanderbilt Law Review* 47(1): 43–113.
- Clermont, Kevin M. (1987). "Procedure's Magical Number Three: Psychological Bases for Standards of Decision." *Cornell Law Review* 72(6): 1115–56.
- Cullison, Alan D. (1969). "Probability Analysis of Judicial Fact-Finding: A Preliminary Outline of the Subjective Approach." *University of Toledo Law Review* 1(3): 538–98.
- Daubert v. Merrell Dow Pharmaceuticals* (1993), 509 U.S. 579.
- Doll, Jen (2012). "New York Does Convict Rapists, but They Don't Make the Front Page." *Atlantic Wire*. <http://www.theatlanticwire.com/national/2012/03/new-york-does-convict-rapists-they-dont-make-front-page/50516/> (visited October 30, 2012).
- Eisenberg, Theodore, Jeffery J. Rachlinski, and Martin T. Wells (2002). "Reconciling Experimental Incoherence with Real-World Punitive Damages." *Stanford Law Review* 54(6): 1239–71.
- Engel, Christoph (2009). "Preponderance of the Evidence versus In-Time Conviction: A Behavioral Perspective on a Conflict between American and Continental European Law." *Vermont Law Review* 33(3): 435–67.
- Fischhoff, Baruch (1975). "Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment under Uncertainty." *Journal of Experimental Psychology: Human Perceptions & Performance* 1(3): 288–99.
- Friedman, Richard D. (2003). "Minimizing the Jury Over-valuation Concern." *Michigan State Law Review* 2003(4): 967–86.
- Froeb, Luke M., and Bruce H. Kobayashi (1996). "Naïve, Biased, Yet Bayesian: Can Juries Interpret Selectively Produced Evidence?" *Journal of Law, Economics & Organization* 12(1): 257–76.
- Gold, Victor J. (1983). "Federal Rule of Evidence 403: Observations on the Nature of Unfairly Prejudicial Evidence." *Washington Law Review* 58(3): 497–533.
- Gold, Victor J. (1987). "Covert Advocacy: Reflections on the Use of Psychological Persuasion Techniques in the Courtroom." *North Carolina Law Review* 65(3): 481–515.
- Goldman, Alvin I. (2003). "Simple Heuristics and Legal Evidence." *Law, Probability & Risk* 2: 215–26.
- Greenfeld, Lawrence A. (1997). *Bureau of Justice Statistics, Sex Offenses and Offenders: An Analysis of Data on Rape and Sexual Assault*. February.
- Guthrie, Chris, Jeffrey J. Rachlinski and Andrew J. Wistrich (2001). "Inside the Judicial Mind." *Cornell Law Review* 86(4): 777–830.
- Guttel, Ehud (2004). "Overcorrection." *Georgetown Law Journal* 93(1): 241–83.
- Guttel, Ehud, and Doron Teichman (2012). "Criminal Sanctions in the Defense of the Innocent." *Michigan Law Review* 110(4): 597–645.



- Hastie, Reid, David A. Schkade, and John W. Payne (1999). "Juror Judgments in Civil Cases: Hindsight Effects on Judgments of Liability for Punitive Damages." *Law & Human Behavior* 23(5): 597–614.
- Heller, Kevin Jon (2006). "The Cognitive Psychology of Circumstantial Evidence." *Michigan Law Review* 105(2): 241–305.
- Hutchins, Robert M., and Donald Slesinger (1928). "Some Observations on the Law of Evidence." *Columbia Law Review* 28(4): 432–40.
- Jolls, Christine, Cass R. Sunstein, and Richard Thaler (1998). "A Behavioral Approach to Law and Economics." *Stanford Law Review* 50(5): 1471–550.
- Kagehiro, Dorothy K. (1990). "Defining the Standard of Proof in Jury Instructions." *Psychological Science* 1(3): 194–200.
- Kahan, Dan M. (2010). "The Economics—Conventional, Behavioral, and Political—of 'Subsequent Remedial Measures' Evidence." *Columbia Law Review* 110(6): 1616–53.
- Kahneman, Daniel and Amos Tversky (1982). "Subjective Probability: A Judgment of Representativeness." In Kahneman, Daniel, Paul Slovic, and Amos Tversky, *Judgement Under Uncertainty: Heuristics and Biases*, 32–47.
- Kaplan, John (1968). "Decision Theory and the Factfinding Process." *Stanford Law Review* 20(6): 1065–92.
- Kaplow, Louis (2012). "Burdens of Proof." *Yale Law Journal* 121(4): 738–859.
- Kaye, David (1979). "Probability Theory Meets Res Ipsa Loquitur." *Michigan Law Review* 77(6): 1456–84.
- Kessler, Elizabeth (1992). "Pattern of Sexual Conduct Evidence and Present Consent: Limiting the Admissibility of Sexual History Evidence in Rape Prosecutions." *Women's Rights Law Reporter* 14(1): 79–96.
- Korobkin, Russell B., and Thomas S. Ulen (2000). "Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics." *California Law Review* 88(4): 1051–144.
- Langan, Patrick A., Erica L. Schmitt, and Matthew R. Durose (2003). *Bureau of Justice Statistics, Recidivism of Sex Offenders Released from Prison in 1994*.
- Leo, Richard A., and Richard J. Ofshe (1998). "The Consequences of False Confessions: Deprivations of Liberty and Miscarriages of Justice in the Age of Psychological Interrogation." *Journal of Criminal Law and Criminology* 88(2): 429–96.
- Levenson, Jill, Yolanda N. Brannon, Timothy Fortney and Juanita Baker (2007). "Public Perceptions about Sex Offenders and Community Protect Policies." *Analyses of Social Issues and Public Policy* 7(1): 137–61.
- Lyon, Thomas D., and Jonathan J. Koehler (1996). "The Relevance Ratio: Evaluating the Probative Value of Expert Testimony in Child Sexual Abuse Cases." *Cornell Law Review* 82(1): 43–78.
- McCauliff, C. M. A. (1982). "Burdens of Proof: Degrees of Belief, Quanta of Evidence or Constitutional Guarantees?" *Vanderbilt Law Review* 35(6): 1293–335.
- Meadow, William, and Cass R. Sunstein (2001). "Statistics, Not Experts." *Duke Law Journal* 51(2): 629–46.
- Mendez, Miguel Angel (1984). "California's New Law on Character Evidence: Evidence Code Section 352 and the Impact of Recent Psychological Studies." *UCLA Law Review* 31(5): 1003–60.
- Mitchell, Gregory (2003). "Mapping Evidence Law." *Michigan State Law Review* 2003(4): 1065–148.

- Moreno, Joëlle Anne (2003). "Translating 'Visions of Rationality' into Specific Legal Reforms." *Michigan State Law Review* 2003(4): 1175–88.
- National Academy of Sciences (2009). Strengthening Forensic Science in the United States: A Path Forward. <http://www.nap.edu/catalog/12589.html>.
- Park, Roger C. (1996). "Character Evidence Issues in the O.J. Simpson Case." *University of Colorado Law Review* 67(4): 747–76.
- Rachlinski, Jeffrey J. (2000). "Heuristics and Biases in the Courts: Ignorance or Adaptation?" *Oregon Law Review* 79(1): 61–102.
- Rachlinski, Jeffery J. (2011). "The Psychological Foundations of Behavioral Law and Economics." *University of Illinois Law Review* 2011(5): 1675–96.
- Risinger, D. Michael (2007). "Innocents Convicted: An Empirically Justified Factual Wrongful Conviction Rate." *Journal of Criminal Law and Criminology* 97(3): 761–806.
- Robbennolt, Jennifer K. (2005). "Evaluating Juries by Comparison to Judges: A Benchmark for Judging?" *Florida State University Law Review* 32(2): 469–509.
- Saks, Michael J., and Robert F. Kidd (1980). "Human Information Processing and Adjudication: Trial by Heuristics." *Law & Society Review* 15(1): 123–60.
- Sanchirico, Christopher William (2001). "Character Evidence and the Object of Trial." *Columbia Law Review* 101(6): 1227–311.
- Sanchirico, Chris William (2003). "Finding Error." *Michigan State Law Review* 2003(4): 1189–203.
- Sanchirico, Chris William (2004). "Evidence, Procedure, and the Upside of Cognitive Error." *Stanford Law Review* 57(2): 291–365.
- Schacter, Daniel L., Robyn Dawes, Larry L. Jacoby, Daniel Kahneman, Richard Lempert, Henry L. Roediger, and Robert Rosenthal (2008). "Policy Forum: Studying Eyewitness Investigations in the Field." *Law & Human Behavior* 32(1): 3–5.
- Seidmann, Daniel J., and Alex Stein (2000). "The Right to Silence Helps the Innocent: A Game-Theoretic Analysis of the Fifth Amendment Privilege." *Harvard Law Review* 114(2): 430–510.
- Simon, Dan (2004). "A Third View of the Black Box: Cognitive Coherence in Legal Decision Making." *University of Chicago Law Review* 71(2): 511–86.
- Simon, Rita James, and Linda Mahan (1971). "Quantifying Burdens of Proof: A View from the Bench, the Jury and the Classroom." *Law and Society Review* 5(3):319–30.
- Skolnick, Jerome H. (1961). "Scientific Theory and Scientific Evidence: An Analysis of Lie-Detection." *Yale Law Journal* 70(5): 694–728.
- Slobogin, Christopher (1999). "Why Liberals Should Chuck the Exclusionary Rule." *University of Illinois Law Review* 1999(2): 363–446.
- Smith, Hubert W. (1942). "Components of Proof in Legal Proceedings." *Yale Law Journal* 51(4): 537–81.
- Tanford, J. Alexander (1990). "The Law and Psychology of Jury Instruction." *Nebraska Law Review* 69(1): 71–111.
- Tillers, Peter (1998). "What Is Wrong with Character Evidence?" *Hastings Law Journal* 49(3): 781–834.
- Tribe, Laurence H. (1971). "Trial by Mathematics: Precision and Ritual in the Legal Process." *Harvard Law Review* 84(6): 1329–93.
- Tversky, Amos, and Daniel Kahneman (1973). "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5(1): 207–33.

- Tversky, Amos, and Daniel Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases." *Science* 185(4157): 1124–31.
- Tversky, Amos, and Daniel Kahneman (1982). "Judgment under Uncertainty: Heuristics and Biases." In Kahneman, Daniel, Paul Slovic, and Amos Tversky, *Judgement Under Uncertainty: Heuristics and Biases*, 3–22.
- Vars, Fredrick E. (2010). "Toward a General Theory of Standards of Proof." *Catholic University Law Review* 60(1): 1–45.
- Wellborn, Olin Guy III (1991). "Demeanor." *Cornell Law Review* 76(5): 1075–105.
- Wistrich, Andrew J., Chris Guthrie, and Jeffrey J. Rachlinski (2005). "Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding." *University of Pennsylvania Law Review* 153(4): 1251–345.
- Woocher, Fredric D. (1977). "Did Your Eyes Deceive You? Expert Psychological Testimony on the Unreliability of Eyewitness Identification." *Stanford Law Review* 29(5): 969–1030.
- Worthington, Debra L., Merrie Jo Stallard, Joseph M. Price, and Peter J. Goss (2002). "Hindsight Bias, Daubert, and the Silicone Breast Implant Litigation." *Psychology Public Policy and Law* 8(2): 154–79.
- Zamir, Eyal, and Ilana Ritov (2012). "Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation." *Journal of Legal Studies* 41(1): 165–207.
- Zamir, Eyal, Illana Ritov, and Doron Teichman (2012). "Seeing Is Believing: The Anti-inference Bias." [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1989561](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1989561) (visited April 21, 2012).

## CHAPTER 28

---

# NUDGES.GOV

## *Behaviorally Informed Regulation*

---

CASS R. SUNSTEIN

### 1 INTRODUCTION

---

IN recent decades, cognitive psychologists and behavioral economists have been incorporating empirical findings about human behavior into economic models. These findings are transforming people's thinking about regulation and its likely consequences. They also offer some suggestions about the appropriate design of “nudges”—low-cost, choice-preserving, behaviorally informed approaches to regulatory problems, including disclosure requirements, default rules, and simplification (Thaler and Sunstein 2008; Sunstein 2013).

A general lesson is that *choice architecture*, understood as the background against which decisions are made, has major consequences for both decisions and outcomes. As a result, small, inexpensive policy initiatives can have large and highly beneficial effects in areas that include health, energy, the environment, savings, and more. The purpose of this chapter is to explore relevant evidence, to catalogue behaviorally informed practices and reforms, and to discuss some implications for regulatory policy. And while the primary focus is on small, inexpensive regulatory initiatives, there is a still more general theme, which involves the importance of ensuring that regulations have strong behavioral foundations.

I write in part on the basis of my experience as administrator of the White House Office of Information and Regulatory Affairs, where I was privileged to serve between 2009 and 2012. In that period, a number of people in the Obama administration took the findings of behavioral economics quite seriously. We adopted a large number of initiatives that count as nudges. One of my main goals here is to catalogue those initiatives and to explore their implications for the future.

In the United States, regulatory efforts have been directly informed by behavioral findings, and behavioral economics has played an unmistakable role in numerous domains. The relevant initiatives enlist tools such as disclosure, warnings, and default rules, and they can be found in multiple areas, including fuel economy, energy efficiency, environmental protection, healthcare, and obesity. As a result, behavioral findings have become an important reference point for regulatory and other policymaking in the United States.

In the United Kingdom, Prime Minister Cameron has created a Behavioural Insights Team with the specific goal of incorporating an understanding of human behavior into policy initiatives. The official website states that its “work draws on insights from the growing body of academic research in the fields of behavioural economics and psychology which show how often subtle changes to the way in which decisions are framed can have big impacts on how people respond to them.” (Cabinet Office, n.d.) The Team has used these insights to promote initiatives in numerous areas, including smoking cessation, energy efficiency, organ donation, consumer protection, and compliance strategies in general. A great deal of money is being saved. Other nations have expressed keen interest in the work of the Team, and its operations are expanding. In 2013, the United States created a behavioral insights team of its own.

Behavioral economics has drawn attention in Europe more broadly. The Organisation for Economic Development and Cooperation (OECD) has published a Consumer Policy Toolkit that recommends a number of initiatives rooted in behavioral findings (OECD 2010). In the European Union, the Directorate-General for Health and Consumers has also shown the influence of behavioral economics (DG SANCO 2010). A report from the European Commission, called *Green Behavior*, enlists behavioral economics to outline policy initiatives to protect the environment (European Commission 2012; [inudgeyou.com](http://inudgeyou.com), n.d.). Private organizations are also using behavioral insights to promote a variety of environmental, health-related, and other goals (see [inudgeyou.com](http://inudgeyou.com), n.d.; [greeNudge.no](http://greeNudge.no)).

It is clear that behavioral findings are having a large impact on regulation, law, and public policy all over the world, and that impact is likely to grow over the next decades. The interest cuts across conventional political categories, and because behavioral findings suggest the possibility of low-cost, high-impact interventions, they are likely to attract considerable attention in economically challenging times. For this reason, it is particularly important to have a sense of what we know, what we do not know, and how emerging understandings can inform sensible policies and reforms.

The structure of this chapter is as follows. It begins with a brief outline of empirical findings of particular relevance to regulatory policy. It then discusses, in sequence, four sets of applications, involving disclosure, default rules, salience, and social norms.

## 2 WHAT WE KNOW

---

### 2.1 Findings

For purposes of regulation, the central findings of behavioral research fall in four categories. What follows is not meant to be a comprehensive account; the focus is on those findings that bear directly on regulatory policy.

#### 2.1.1 *Inertia and Procrastination*

##### 2.1.1.1. *Default rules often have a large effect on social outcomes.*

Both private and public institutions often establish “default rules”—rules that determine the result if people make no affirmative choice at all. In part because of the power of inertia, default rules can be extremely important. In the domain of retirement savings, for example, the default rule has significant consequences. When people are asked whether they want to opt in to a retirement plan, the level of participation is far lower than if they are asked whether they want to opt out. Automatic enrollment significantly increases participation (for further details and citations see chapter 11 by Zamir and chapter 12 by Korobkin in this volume).

More generally, people may decline to change from the status quo even if the costs of change are low and the benefits substantial. In the context of energy and the environment, for example, we might predict that people would neglect to switch to fuel-efficient alternatives even when it is in their interest to do so. It follows that complexity can have serious adverse effects, by increasing the power of inertia, and that ease and simplification (including reduction of paperwork burdens) can produce significant benefits. These benefits include increased compliance with law and greater participation in public programs (see Sunstein 2011 and Sunstein 2013, for examples).

##### 2.1.1.2. *Procrastination can have significant adverse effects.*

According to standard economic theory, people will consider both the short term and the long term. They will take account of relevant uncertainties; the future may be unpredictable, and significant changes may occur over time. They will appropriately discount the future. In practice, however, some people procrastinate or neglect to take steps that impose small short-term costs but that would produce large long-term gains. They may, for example, delay enrolling in a retirement plan, starting to exercise, ceasing to smoke, or using some valuable, cost-saving technology. When procrastination

is creating significant problems, automatic enrollment in relevant programs might be helpful.

*2.1.1.3. When people are informed of the benefits or risks of engaging in certain actions, they are far more likely to act in accordance with that information if they are simultaneously provided with clear, explicit information about how to do so.*

For example, those who are informed of the benefits of a vaccine are more likely to become vaccinated if they are also given specific plans and maps describing where to go (Leventhal, Singer, and Jones 1965; Nickerson and Rogers 2010). Similarly, behavior has been shown to be significantly affected if people are informed, not abstractly of the value of “healthy eating,” but specifically of the advantages of buying 1% milk (as opposed to whole milk) (Heath and Heath 2010). In many domains, the identification of a specific, clear, unambiguous path or plan has an important effect on social outcomes; complexity or vagueness can ensure inaction, even when people are informed about risks and potential improvements.

### *2.1.2 Framing and Presentation*

*2.1.2.1. People can be influenced by how information is presented or “framed.”*

If, for example, people are informed that they will *gain* a certain amount of money by using energy-efficient products, they may be less likely to change their behavior than if they are told that they will *lose* the same amount of money by not using such products (Gonzales, Aronson, and Costanzo 1988). It follows that a product that is labeled “90 percent fat-free” may well be more appealing than one that is labeled “10 percent fat.” It also follows that choices are often not made based solely on their consequences; assessments may be affected by the relevant frame (Redelmeier, Rozin, and Kahneman 1993; Levin, Schneider, and Gaeth 1998).

*2.1.2.2. Information that is vivid and salient can have a larger impact on behavior than information that is statistical and abstract.*

With respect to public health, it is well-known that vivid displays can be more effective than abstract presentations of statistical risks. This point bears on the design of effective warnings. Attention is a scarce resource, and vivid, salient, and novel presentations may trigger attention in ways that abstract or familiar ones cannot.

In particular, salience greatly matters. Why, for example, do people pay bank overdraft fees? One of the many possible answers is that such fees are not sufficiently salient to people, and the fees are incurred as a result of inattention or inadvertent mistakes. One study suggests that limited attention is indeed a source of the problem, and that once overdraft fees become salient, they are significantly reduced (Stango and Zinman 2011). When people take surveys about such fees, they are less likely to incur a fee in



the following month, and when they take a number of surveys, the issue becomes sufficiently salient that overdraft fees are reduced for as much as two years. In many areas, the mere act of being surveyed can affect behavior by, for example, increasing use of water treatment products (thus promoting health) and the take-up of health insurance; one reason is that being surveyed increases the salience of the action in question (Zwane et al. 2011).

A more general point is that many costs (or benefits) are less salient than purchase prices; they are “shrouded attributes” to which some consumers do not pay much attention. Such “add-on” costs may matter a great deal but receive little consideration, because they are not salient (see chapter 18 by Bar-Gill in this volume).

### *2.1.2.3 People often display loss aversion; they may well dislike losses more than they like corresponding gains.*

Whether a change counts as a loss or a gain depends on the *reference point*, which can be affected by policy decisions, and which is often the status quo (Kahneman and Tversky 1979; for further details and citations see chapter 11 by Zamir in this volume). In part for this reason, the initial allocation of a legal entitlement can affect people’s valuations; those who have the initial allocation may value a good more than they would if the allocation were originally elsewhere, thus showing an *endowment effect* (see chapter 12 by Korobkin in this volume).

## *2.1.3 Social Influences*

### *2.1.3.1. In multiple domains, individual behavior is greatly influenced by the perceived behavior of other people*

With respect to obesity, proper exercise, alcohol consumption, smoking, becoming vaccinated, and much more, the perceived decisions of others have a significant influence on individual behavior and choice (Hirshleifer 1995; Duflo and Saez 2003). The behavior of peers has been found to have a significant effect on risky behavior among adolescents, including tobacco smoking, marijuana use, and truancy (Card and Guiliano 2011; Bisin, Moro, and Topa 2011).

In particular, food consumption is greatly affected by the food consumption of others, and indeed the body type of others in the relevant group can affect people’s responses to their food choices, with a greater effect from those who are thin than those who are heavy (McFerran et al. 2011). The norm conveys significant information about what ought to be done; for that reason, those who lack private information may follow the apparent beliefs and behavior of relevant others, sometimes creating *informational cascades*.

In addition, people care about their reputations, and for that reason, they may be influenced by others so as not to incur their disapproval. In some contexts, social norms can help create a phenomenon of *compliance without enforcement*—as, for example, when people comply with laws forbidding indoor smoking or requiring buckling

of seat belts, in part because of social norms or the expressive function of those laws. These points bear on the value and importance, in many domains, of private-public partnerships.

*2.1.3.2. People are more likely to cooperate with one another, and to contribute to the solution of collective action problems, than standard economic theory predicts.*

People's willingness to cooperate is partly a product of an independent commitment to fairness; but it is partly a product of a belief that others will see and punish a failure to cooperate or to act fairly. Norms of reciprocity can be exceedingly important (Camerer 2003).

#### *2.1.4 Probability Assessment and Attitude to Risk*

*2.1.4.1. In some domains, people show unrealistic optimism*

The "above average" effect is common (Weinstein 1987); many people believe that they are less likely than others to suffer from various misfortunes, including automobile accidents and adverse health outcomes (Jolls 1998; Sharot 2011; chapter 13 by Williams in this volume). One study found that while smokers do not underestimate statistical risks faced by the population of smokers, they nonetheless believe that their personal risk is less than that of the average smoker (Slovic 1998).

*2.1.4.2. People often use heuristics, or mental shortcuts, when assessing risks.*

For example, judgments about probability are often affected by whether a recent event comes readily to mind (Kahneman and Frederick 2002; Tversky and Kahneman 1973). If an event is cognitively "available," people may well overestimate its probability, and vice versa (see generally chapter 13 by Williams in this volume).

*2.1.4.3. People sometimes do not make judgments on the basis of expected value, and they may neglect or disregard the issue of probability, especially when strong emotions are triggered.*

When emotions are strongly felt, people may focus on the outcome and not on the probability that it will occur (Rottenstreich and Hsee 2001). This point obviously bears on reactions to extreme events of various sorts.

*2.1.4.4. People's attitudes to risk are reference-dependent.*

Prospect theory suggests that for moderate to high probability changes, people tend to be risk averse with respect to gains but risk seeking with respect to losses. For very small probabilities, people tend to be risk seeking with respect to gains but risk averse for losses (hence gambling and insurance) (Kahneman and Tversky 1979).

### 2.1.5 *Summary*

These various findings are hardly inconsistent with the conventional economic emphasis on the importance of material incentives; actual and perceived costs and benefits certainly matter. When the price of a product rises, or when it becomes clear that use of a product imposes serious health risks, the demand for the product is likely to fall (at least, and this is a significant qualification, if these effects are salient). But apart from strictly material incentives of this kind, evidence suggests the independent importance of (1) the choice architecture and (2) prevailing social norms. If, for example, healthy foods are prominent and easily accessible, people are more likely to choose them; one study finds an 8% to 16% decrease in intake simply by making food more difficult to reach (as, for example, by varying its proximity by ten inches or altering the serving utensil) (Rozin et al. 2011). The problem of childhood obesity is, at least in part, a result of the easy availability of unhealthy foods. The same point bears on smoking and alcohol abuse. And with respect to savings behavior, automatic enrollment might have larger effects than even significant economic incentives (Chetty et al. 2012).

Here is another way to put the point. The existing social environment and current social norms provide the backdrop for many outcomes. Consumer products are accompanied by default rules of various sorts; consider, for example, rental car and cell phone agreements, where it is possible to opt in or to opt out of a range of features, and where the default rule may much matter. With respect to water quality, air quality, sewage treatment, immunization, and healthcare, the social environment provides relevant background, which is often taken for granted, and which need not, for many people much of the time, become a serious source of deliberation and choice. In particular for people who are well-off, the relevant background, which need not be an object of reflection, is highly desirable and may be taken for granted without causing harm. For others, the background is not so benign, and it should in any case be an object of reflection and choice (Sunstein 2013).

## 2.2 Concerns

### 2.2.1 *Are Predictions Possible?*

It is tempting to respond that these diverse findings might point in different directions, even for the same subpopulation faced with the same problem, and hence that clear predictions cannot be made in particular cases (cf. chapter 7 by Mitchell in this volume). For example, will people save too little or too much? Will they take optimal, excessive, or insufficient precautions against the risks associated with poor diet?

By itself and in the abstract, an understanding of loss aversion, the availability heuristic, and social influences does not produce clear answers. Such an understanding could, on plausible assumptions, suggest that people may save too much

or take excessive precautions, or on other plausible assumptions, suggest the opposite conclusions. And it may well be the case that loss aversion, unrealistic optimism, the availability heuristic, and social influences are simultaneously at work and will point in different directions, making predictions difficult or impossible. For example, unrealistic optimism may lead people to underestimate certain risks, while the availability heuristic may lead people to overestimate the same risks.

It is true that if these findings are taken as a whole and in the abstract, they will not lead to a clear or unique prediction about behavior. Particular situations must be investigated in detail in order to understand likely outcomes. For the purposes of this chapter, however, it is not necessary to engage these questions in detail. Low-cost regulatory policies, such as disclosure and simplification, may be justified even if we do not have a clear understanding, in the abstract, of whether relevant behavior is affected by loss aversion or social influences. Of course it is also true that the design of a disclosure policy should be based on an understanding of how people process information, and that a sensible approach to simplification will require an understanding of whether and why complexity can create problems and of what kinds of simplification can eliminate those problems.

### 2.2.2 *Markets, Government, and the Vexing Problem of Paternalism*

It is natural to wonder whether an understanding of the findings outlined above justifies paternalism, or operates as a defense of “more” regulation. With respect to paternalism in particular, it is true that some of the relevant findings supplement the standard accounts of market failures, suggesting that in some settings, markets may fail, in the sense that they may not promote social welfare even in the presence of perfect competition and full information. We are now in a position to identify a series of behavioral market failures, and these do appear to justify regulatory controls. Responses to behavioral market failures might be counted as paternalistic.

If, for example, people focus on short-term costs and neglect long-term benefits, it is possible that disclosure policies that specifically emphasize the long-term, or even regulatory requirements (involving, for example, energy efficiency), may be justified. It is also possible to identify “internalities”—problems of self-control and errors in judgments that produce within-person harms, as, for example, when smoking behavior leads to serious risks because of the victory of short-term considerations over the longer view. These too count as behavioral market failures, and responses may be paternalistic in character. Richard Thaler and I have argued in defense of “libertarian paternalism” (Thaler and Sunstein 2008; see also Sunstein 2013), understood as approaches that preserve freedom of choice while also steering people in directions that will make their lives go better (by their own lights). And it would be possible to think that at least some behavioral market failures justify more coercive forms of paternalism (more coercive paternalistic measures are advocated on efficiency grounds by Zamir [1998]).

But even if the standard accounts of potential market failures are supplemented, it does not necessarily follow that paternalism, or more regulation, is justified. Perhaps

markets will eventually address the problem better than regulators would, and for multiple reasons, the cure might be worse than the disease.

Indeed, some of the findings might argue in favor of less rather than more regulation and less rather than more paternalism. Market forces can provide a great deal of help in the face of human error. For example, the private sector has relied increasingly on automatic enrollment in savings plans, and countless companies attempt to promote better diet and more exercise (often expecting to obtain more customers as a result).

It should not be necessary to emphasize that public officials are subject to error as well. Indeed, errors may result from one or more of the findings traced above; officials are human and capable of error too. The dynamics of the political process may or may not lead in the right direction. It would be absurd to say that behaviorally informed regulation is more aggressive than regulation that is not so informed, or that an understanding of recent empirical findings calls for more regulation rather than less. The argument is instead that such an understanding can help to inform the design of regulatory programs.

With respect to the particular concerns, it would be valuable to have a better understanding of how the relevant findings apply within heterogeneous groups; the findings are far from uniform within the population, and for purposes of policy, heterogeneity may matter. It would also be valuable to have a better understanding of actual conduct within diverse settings. For example, we have good reason to believe that many people do not buy energy-efficient products even when it would be in their economic interest to do so, but the conceptual and empirical issues are complex and have not been fully sorted out (cf. Alcott 2011).

But even at this stage, existing research offers helpful lessons for regulatory policy. Relevant research suggests that four such approaches have particular promise: (1) using disclosure as a regulatory tool, especially if disclosure policies are designed with an appreciation of how people process information; (2) simplifying and easing choices through appropriate default rules, reduction of complexity and paperwork requirements, and related strategies; (3) increasing the salience of certain factors or variables; and (4) promoting social norms through private-public partnerships and other approaches that operate in the service of agreed-upon public goals. Behaviorally informed approaches of this kind are already in place, including a number of recent initiatives.

---

### 3. DISCLOSURE AS A NUDGE

---

In this section, I explore the uses of disclosure as a behaviorally informed regulatory tool. It is important to distinguish between *summary disclosure*, often provided at the point of purchase, and *full disclosure*, typically provided on the Internet. A central point is that disclosure policies should be based on an understanding of how people process information.

## 3.1 Actually Informing Choice

### 3.1.1 *Examples*

Many statutory programs recognize that information disclosure can be a useful regulatory tool, replacing or complementing other approaches. Central examples include legislative efforts to require disclosure of the risks associated with smoking, of potential savings from energy efficiency, and of information that bears on health. Recent initiatives have drawn directly from behavioral economics, emphasizing the importance of plain language, clarity, and simplicity.

#### 3.1.1.1 *Nutrition*

In the domain of nutrition, a number of disclosure requirements are in force. To take just one example, a final rule has been issued by the US Department of Agriculture (USDA), requiring provision of nutritional information to consumers with respect to meat and poultry products. Nutrition Facts panels must be provided on the labels of such products. Under the rule, the panels must contain information with respect to calories and both total and saturated fats (9 CFR § 317.309).

The rule clearly recognizes the potential importance of framing. If a product lists a percentage statement such as “80% lean,” it must also list its fat percentage. This requirement should avoid the confusion that can result from selective framing; a statement that a product is 80% lean, standing by itself, makes leanness salient, and may therefore be misleading.

#### 3.1.1.2 *Healthcare*

The Patient Protection and Affordable Care Act of 2010 (Affordable Care Act) contains a large number of disclosure requirements designed to promote accountability and informed choice with respect to healthcare. Indeed, the Affordable Care Act is, in significant part, a series of disclosure requirements, many of which are meant to inform consumers, and to do so in a way that is alert to behavioral findings. Under the act, a restaurant that is part of a chain with twenty or more locations doing business under the same name is required to disclose calories on the menu board. Such restaurants are also required to provide in a written form (available to customers upon request) additional nutrition information pertaining to total calories and calories from fat, as well as amounts of fat, saturated fat, cholesterol, sodium, total carbohydrates, and more. (Affordable Care Act 2010). There continues, of course, to be a dispute about the actual effects of disclosure requirements of this kind, and further evidence is indispensable (for a finding of positive results, in the sense of decreased caloric intake, see Bollinger, Leslie, and Sorenson 2010).

In a similar vein, § 1103 of the act calls for “[i]mmediate information that allows consumers to identify affordable coverage options.” It requires the establishment of an Internet portal to enable beneficiaries to find affordable and comprehensive coverage options, including information about eligibility, availability, premium rates, cost sharing, and the percentage of total premium revenues spent on healthcare, rather than administrative expenses.

It should be clear from this brief survey that the range of recent disclosure requirements is very wide. If carefully designed, such approaches have considerable promise.

### 3.1.2 *How, Not Only Whether*

As social scientists have emphasized, disclosure as such may not be enough; regulators should devote care and attention to *how, not only whether, disclosure occurs*. Clarity and simplicity are often critical (Sunstein 2011). In some cases, accurate disclosure of information may be ineffective if the information is too abstract, vague, detailed, complex, poorly framed, or overwhelming to be useful. Disclosure requirements should be designed for *homo sapiens*, not *homo economicus*. In addition, emphasis on certain variables may attract undue attention and prove to be misleading. If disclosure requirements are to be helpful, they must be designed to be sensitive to how people actually process information.

A good rule of thumb is that disclosure should be concrete, straightforward, simple, meaningful, timely, and salient. If the goal is to inform people about how to avoid risks or to obtain benefits, disclosure should avoid abstract statements (involving, for example, “healthy eating”) and instead clearly identify the steps that might be taken to obtain the relevant goal (by specifying, for example, what specific actions parents might take to reduce the risk of childhood obesity).

### 3.1.3 *Testing Disclosure*

To the extent possible, agencies should study in advance the actual effects of alternative disclosure designs to ensure that information is properly presented and will actually inform consumers. The Nutrition Facts labels on many food products followed such a process of advance study, with careful investigation of consumer responses to different presentations of the relevant material. Actual experience in the marketplace can, of course, provide valuable information.

Because they are more likely to yield information about actual behavior, experimental or quasi-experimental studies are preferred to focus groups, which generally ask people, under highly artificial conditions, what they think and which may therefore fail to capture actual behavior. Randomized field experiments have particular advantages (Greenstone 2009; see also chapter 5 by Engel in this volume). At the same time, focus groups can be useful, especially if they are carefully designed to assess likely behavior (rather than simply asking people which presentations or formats they most like).



### 3.1.4 *Avoiding Confusion*

If not carefully designed, disclosure requirements can produce ineffective, confusing, and potentially misleading messages. Behaviorally informed approaches are alert to this risk and suggest possible improvements.

For instance, automobile manufacturers are currently required to disclose the fuel economy of new vehicles as measured by miles per gallon (MPG). This disclosure is useful for consumers and helps to promote informed choice. As the Environmental Protection Agency (EPA) has emphasized, however, MPG is a nonlinear measure of fuel consumption (Environmental Protection Agency 2009a). For a fixed travel distance, a change from 20 to 25 MPG produces a larger reduction in fuel costs than does a change from 30 to 35 MPG, or even from 30 to 38 MPG. To see the point more dramatically, consider the fact that an increase from 10 to 20 MPG (which reduces the consumption of fuel from 10 to 5 gallons per 100 miles) produces more savings than an increase from 20 to 40 MPG (which reduces fuel consumption from 5 to 2.5 gallons per 100 miles).

Evidence suggests that many consumers do not understand this point and tend to interpret MPG as linear with fuel costs. When it occurs, this error is likely to produce inadequately informed purchasing decisions when people are making comparative judgments about fuel costs. Consumers tend to *underestimate* the cost differences among low-MPG vehicles and tend to *overestimate* the cost differences among high-MPG vehicles (Allcott 2011). By contrast, an alternative fuel economy metric, such as gallons per mile, could be far less confusing. Such a measure is linear with fuel costs and hence suggests a possible way to help consumers make better choices. Recognizing the imperfections and potentially misleading nature of the MPG measure, the Department of Transportation and EPA mandated a label including a clear statement about anticipated fuel savings (or costs) over a five-year period.

In a related vein, the USDA has abandoned the “Food Pyramid,” used for years as the central icon to promote healthy eating. The Pyramid has long been criticized as insufficiently informative; it does not offer people any kind of clear “path” with respect to healthy diet and does not connect to people’s actual experience with food (Heath and Heath 2010). In response, the USDA replaced the Pyramid with a new, simpler icon, consisting of a plate with clear markings for fruit, vegetable, grains, and protein. The plate is accompanied by straightforward guidance, including “make half your plate fruits and vegetables,” “drink water instead of sugary drinks,” and “switch to fat-free or low-fat (1%) milk.” This approach has the key advantage of informing people what to do, if they seek to have a healthier diet.

In a related vein, the HHS, implementing a provision of the Affordable Care Act, has finalized a rule to require insurance companies to provide clear, plain language summaries of relevant information to prospective customers. The rule includes basic information, including the annual premium, the annual deductible, a statement of services that are not covered, and a statement of costs for going to an out-of-network provider.

In some circumstances, the tendency toward unrealistic optimism may lead some consumers to downplay or neglect information about statistical risks associated with a product or an activity. Possible examples include smoking and distracted driving. In such circumstances, disclosure might be designed to make the risks associated with the product less abstract, more vivid, and salient. For example, the Family Smoking Prevention and Tobacco Control Act of 2009 (Smoking Prevention Act) requires graphic warnings with respect to the risks of smoking tobacco, and in 2011 the Food and Drug Administration (FDA) finalized such warnings, with vivid and even disturbing pictures of some of the adverse outcomes associated with smoking. The warnings were later invalidated in court.

### 3.1.5 *Promoting Competition*

If disclosure requirements are straightforward and simple, they should facilitate comparison shopping and hence market competition. Drawing on social science research (for overviews, see Thaler and Sunstein 2008; Sunstein 2013), the Treasury Department's account of financial regulation emphasizes the value of requiring that "communications with the consumer are reasonable, not merely technically compliant and non-deceptive. Reasonableness includes balance in the presentation of risks and benefits, as well as clarity and conspicuousness in the description of significant product costs and risks." (Department of the Treasury 2009). The department's analysis goes on to say that one goal should be to

harness technology to make disclosures more dynamic and adaptable to the needs of the individual consumer. . . . Disclosures should show consumers the consequences of their financial decisions. . . . [The regulator] should. . . mandate or encourage calculator disclosures for mortgages to assist with comparison shopping. For example, a calculator that shows the costs of a mortgage based on the consumer's expectations for how long she will stay in the home may reveal a more significant difference between two products than appears on standard paper disclosures. (Department of the Treasury 2009)

In keeping with this theme, the Consumer Financial Protection Bureau is authorized to ensure that "consumers are provided with timely and understandable information to make responsible decisions about financial transactions" (Dodd-Frank Act 2010). The Bureau is also authorized to issue rules that ensure that information is "fully, accurately, and effectively disclosed to consumers in a manner that permits consumers to understand the costs, benefits, and risks associated with the product or service, in light of the facts and circumstances" (Dodd-Frank Act 2010).

To accomplish this task, the Bureau is authorized to issue model forms with "a clear and conspicuous disclosure that, at a minimum—(A) uses plain language comprehensible to consumers; (B) contains a clear format and design, such as an easily readable type font; and (C) succinctly explains the information that must be communicated to the consumer" (Dodd-Frank Act 2010; Riis and Ratner 2010). In addition, the director of the Bureau is required to "establish a unit whose functions shall include researching, analyzing, and reporting on. . . consumer awareness, understanding, and use of

disclosures and communications regarding consumer financial products or services” and “consumer behavior with respect to consumer financial products or services, including performance on mortgage loans.” Note that new technologies make it possible to inform consumers of their own choices and usages, an approach that may be especially important when firms have better information than consumers do about such choices and usages (see chapter 18 by Bar-Gill in this volume).

In the same general vein, the Department of Labor issued a final rule requiring disclosure to workers of relevant information in pension plans. The rule is designed to require clear, simple disclosure of information about fees and expenses and to allow meaningful comparisons, in part through the use of standard methodologies in the calculation and disclosure of expense and return information (29 CFR § 2550.404a-5).

Yet another example is provided by a final rule of the Department of Education that promotes transparency and consumer choice with respect to for-profit education by requiring institutions to provide clear disclosure of costs, debt levels, graduation rates, and placement rates (Department of Education 2010a). These disclosures must be included “in promotional materials [the institution] makes available to prospective students” and be “[p]rominently provide[d] . . . in a simple and meaningful manner on the home page of its program Web site” (34 CFR § 668.6); Department of Education 2010b).

### 3.2 Summary Disclosure and Full Disclosure

Disclosure requirements of this kind are designed to inform consumers at the point of purchase, often with brief summaries of relevant information. Such “summary disclosures” are often complemented with more robust information, typically found on public or private websites. For example, the EPA offers a great deal of material on fuel economy online, going well beyond the information that is available on stickers, and the nutrition facts label is supplemented by a great deal of nutritional information on government websites. Approaches of this kind provide information that private individuals and institutions can adapt, reassemble, and present in new, helpful, imaginative, and often unanticipated ways. Some of the most valuable and creative uses of full disclosure are made by the private sector.

Other disclosure requirements are not specifically directed to consumers or end users at all. They promote public understanding of existing problems and help produce possible solutions by informing people about current practices. One example is the Emergency Planning and Community Right-to-Know Act of 1986 (Emergency Planning Act). At first, this law seemed to be largely a bookkeeping measure, requiring a “Toxic Release Inventory” in which firms reported what pollutants they were releasing. But available evidence indicates that it has had beneficial effects, helping to spur reductions in toxic releases throughout the United States (Hamilton 2005). One reason involves public accountability: public attention can help promote behavior that fits with statutory purposes.

In 2009 and 2010, the Occupational Safety and Health Administration (OSHA) placed a significant subset of its fatality, illness, and injury data online, in a step that should promote both accountability and safer workplaces (Department of Labor 2011). In 2009, the EPA issued a greenhouse gas reporting rule, requiring disclosure by many of the most significant emitters (Environmental Protection Agency 2009b). The data may well allow businesses to find innovative ways to track their own emissions, to compare them to similar facilities, and eventually to identify low-cost reductions.

To be sure, mandatory disclosure can impose costs and burdens on both private and public institutions, and to the extent permitted by law, those costs and burdens should be considered when deciding whether and how to proceed. Empirical evidence on the actual effects of disclosure policies is indispensable (Greenstone 2009; Schwartz et al. 2011; Sunstein 2010a).

## 4 DEFAULT RULES AS NUDGES

---

Social science research provides strong evidence that starting points, or default rules, greatly affect social outcomes (chapter 11 by Zamir and chapter 12 by Korobkin in this volume). In some contexts, it may be possible to promote statutory goals with sensible default rules that preserve freedom of choice and that might help to avoid the rigidity, cost, and unintended adverse consequences of mandates and bans. Default rules are one way of easing people's choices, and they are used in countless domains by both public and private institutions.

### 4.1 Automatic Enrollment and Default Rules: Examples

#### 4.1.1 *Savings*

In the United States, employers have long asked workers whether they want to enroll in 401(k) plans; under a common approach, the default rule is nonenrollment. Even when enrollment is easy, the number of employees who enroll, or opt in, has sometimes been relatively low (Gale, Iwry, and Walters 2009). A number of employers have responded by changing the default to automatic enrollment, by which employees are enrolled unless they opt out. The results are clear: significantly more employees end up enrolled with an opt-out design than with opt-in (Gale, Iwry, and Walters 2009). This is so even when opting out is easy. Importantly, automatic enrollment has significant benefits for all groups, with increased anticipated savings for Hispanics, African Americans, and women in particular (Orszag and Rodriguez 2009; Papke, Walker, and Dworsky 2009; Chiteji and Walker 2009).

The Pension Protection Act of 2006 (PPA) (Pension Protection Act 2006) draws directly on these findings by encouraging employers to adopt automatic enrollment plans. The PPA does this by providing nondiscrimination safe harbors for elective deferrals and for matching contributions under plans that include an automatic enrollment feature, as well as by providing protections from state payroll-withholding laws to allow for automatic enrollment. Building on these efforts, President Obama has asked the IRS and the Treasury Department to undertake initiatives to make it easier for employers to adopt such plans (Obama 2009; Internal Revenue Service 2009).

#### 4.1.2 *Healthcare*

A provision of the Affordable Care Act requires employers with over two hundred employees automatically to enroll employees in healthcare plans, while also allowing employees to opt out (Affordable Care Act, 2010). Another provision of the act is called the Community Living Assistance Services and Supports Act (CLASS Act) (CLASS Act, 2010); this provision creates a national voluntary long-term insurance program. The act provides for an automatic enrollment system, whereby employers enroll employees in the program unless they opt out (CLASS Act, 2010). In addition, the act contains an automatic payroll deduction system for the payment of premiums (CLASS Act, 2010).

On February 4, 2010, the Center of Medicare and Medicaid Services (CMS) provided guidance to states via a State Health Official (SHO) letter (Centers for Medicare and Medicaid Services, 2010). In cases where states are able to obtain all the information necessary to determine eligibility, the new option permits states automatically to enroll and renew eligible children in Medicaid or Children's Health Insurance Program (CHIP). This approach allows states to initiate and determine eligibility for Medicaid or CHIP without a signed Medicaid or CHIP program application, as long as the family or child consents to be enrolled in Medicaid or CHIP.

#### 4.1.3 *School Meals*

The National School Lunch Act (Healthy, Hunger-Free Kids Act 2012) takes steps to allow "direct certification" of eligibility, thus reducing complexity and introducing what is a form of automatic enrollment. Under the program, children who are eligible for benefits under certain programs will be "directly eligible" for free lunches and free breakfasts, and hence will not have to fill out additional applications (Healthy, Hunger-Free Kids Act 2012). To promote direct certification, the USDA has issued an interim final rule that is expected to provide up to 270,000 children with school meals (Department of Agriculture 2011) is the total number of directly certified children is now in the millions.

#### 4.1.4 *Payroll Statements*

The Department of Homeland Security has changed the default setting for payroll statements to electronic from paper, thus reducing costs (Orszag 2010). In general, changes of this kind may save significant sums of money for both private and public sectors.

### 4.1.5 *Childhood Obesity*

A great deal of empirical work identifies a noteworthy contributor to the problem of obesity, including childhood obesity. If healthy foods are easily accessible, people are far more likely to choose them, and the same is true for unhealthy foods. Indeed, convenience and accessibility can significantly increase caloric intake (Rozin et al. 2011; Wansink, Just, and McKendry 2010). Some studies have found that when fast-food restaurants are located near schools or residences, significant weight gain occurs in both children and pregnant women (Currie et al. 2010). Even small differences have large effects on food choices and consumption. For example, the sizes of plates and portions have been increasing over time, and they affect how much people eat; when unhealthy foods are made slightly less accessible, their consumption is reduced (Rozin et al. 2011; Wansink 2010; Dayan and Bar-Hillel 2011). These and related issues are discussed in the report of the White House Task Force on Childhood Obesity, which places a great emphasis on the importance of accessibility (White House Task Force on Childhood Obesity 2010).

In a sense, social settings produce something akin to default rules for food choices. These findings—about the importance of seemingly small features of context—have implications for continuing efforts to reduce childhood obesity and many other problems. One study, for example, finds that if people are prompted to consider whether to “downsize” their meals through a simple question, they will eat significantly less at fast-food restaurants (Schwartz et al. 2011). Indeed, the effect of this prompt was found to be greater than that of calorie labeling.

## 4.2 Automatic Enrollment and Default Rules: Mechanisms and Complexities

A great deal of research has attempted to explore exactly why default rules have such a large effect on outcomes (Gale, Iwry, and Walters 2009; Dinner et al. 2009; Carroll et al. 2009). There appear to be three contributing factors. The first involves *inertia and procrastination*. To alter the effect of the default rule, people must make an active choice to reject the default. In view of the power of inertia and the tendency to procrastinate, people may simply continue with the status quo. It follows that self-consciously and well-chosen default rules by individuals, or by private or public institutions, can operate as commitment devices; consider, for example, a default rule in favor of monthly transfer of money into a savings account, or in favor of savings for retirement.

The second factor involves what might be taken to be an *implicit endorsement* of the default rule. Many people appear to conclude that the default was chosen for a reason; they believe that they should not depart from it unless they have particular information to justify a change.

Third, the default rule might establish the *reference point* for people's decisions; the established reference point has significant effects because people dislike losses from that reference point. If, for example, the default rule favors energy-efficient light bulbs, then the loss (in terms of reduced efficiency) may loom large, and there will be a tendency to continue with energy-efficient light bulbs. But if the default rule favors less efficient (and initially less expensive) light bulbs, then the loss in terms of upfront costs may loom large, and there will be a tendency to favor less efficient light bulbs (see also chapter 11 by Zamir and chapter 12 by Korobkin in this volume).

In a significant number of domains, it might be possible to achieve regulatory goals, and to do so while maintaining freedom of choice and at low cost, by selecting good default rules and by avoiding harmful ones. The initial task, of course, is to identify the requirements of the law. Within the context of such requirements, one approach is to select the default rule that reflects what most people would choose if they were adequately informed. Suppose, for example, that a particular default rule would place a strong majority of the relevant population in the situation that they would favor if they made an informed choice. If so, there is a legitimate reason to adopt that default rule (with the understanding that for those who differ from the majority, it remains possible to opt out).

Of course, it may be necessary to do a great deal of work in order to identify the approach that informed people would choose, and on this count, actual evidence about informed choice is extremely important. The issue is simplified if the law requires a particular set of outcomes. A default rule might well make sense if it promotes automatic compliance with the law. Hence it is important to see that use of default rules may serve either as an independent approach, used instead of a mandate or a ban, or as a complementary approach, operating to facilitate compliance with statutory or regulatory requirements.

It is also important to see that default rules can be badly chosen or misused by private and public institutions alike and that some such rules can be harmful. Defaults may have either desirable or undesirable distributional effects. If, for example, poor or poorly educated people are less likely to opt out, defaults may effectively mandate certain outcomes for such people, perhaps to their detriment. The FTC has expressed serious concerns about "negative option marketing," which occurs when those who accept a "free" product are automatically enrolled in a plan or program that carries a monthly fee (unless they explicitly opt out) (16 CFR § 425; Federal Trade Commission 2009). In some cases, negative option marketing has the unfortunate effect of using a default rule to exploit the tendency toward inertia in a way that is harmful to people's welfare; it is easy to imagine both private and public analogues (consider, for example, an automatic enrollment policy that puts an unreasonably large amount of salary into savings).

To evaluate the use of automatic enrollment, the particular circumstances certainly matter. If automatic enrollment is not made transparent to those who are enrolled, it can be considered a form of manipulation, and the problem is worse if it is not in their long-term interest.



Some default rules apply to all of the relevant population, subject to the ability to opt out. Other default rules are *personalized*, in the sense that they draw on available information about which approach best suits individuals in the relevant population. A personalized default might be based on geographical or demographic variables; for example, income and age might be used in determining appropriate default rules for retirement plans. Alternatively, a personalized default might be based on people's own past choices to the extent that they are available (unless those choices are based on an absence of information, manipulation, or some kind of bias).

An advantage of personalized default rules is that they may well be more accurate than "mass" default rules. As technology evolves, it should be increasingly possible to produce personalized defaults, based on people's own choices and situations, and likely to be far more accurate than more general ones. There will be excellent opportunities to use default rules to promote people's welfare. To be sure, any such rules must respect the applicable laws, policies, and regulations involving personal privacy and should avoid unduly crude proxies.

It is important to note that default rules may not "stick" when the relevant population has strong contrary preferences. For example, a study in the United Kingdom found that most people rejected a savings plan with an unusually high default contribution rate (12% of before-tax income) (Beshears et al. 2010). Only about 25% of employees remained at that rate after a year, whereas about 60% of employees remained at a lower default contribution rate. One implication is that "extreme" defaults are less likely to stick; another implication, based on the lower incomes of those who stayed with the default, is that default rules may be more influential for low-income workers than for their higher-earning counterparts (Beshears et al. 2010).

A related finding is that workers were not much affected by a default allocation of a fraction of their tax refund to US savings bonds, apparently because such workers had definite plans to spend their refunds (Bronchetti et al. 2011). A general lesson is that default rules will have a weaker effect, and potentially no effect, when the relevant population has a strong preference for a certain outcome.

### 4.3 Active Choices

An alternative approach, sometimes worth serious consideration, is to avoid any default rule and *to require active choices* (Carroll et al. 2009). Under this approach, people are required to make an actual choice among the various options; they are not defaulted into any particular alternative. With respect to savings, for example, an employer might reject both opt-out and opt-in and simply require employees to indicate their preferences. Evidence suggests that active choices result in far higher levels of savings than a default rules that requires people explicitly to opt in (Carroll et al. 2009).

If inertia and procrastination are playing a significant role, then active choosing may be better than opt-in, in which people end up with outcomes that they would not prefer

if they were to make a choice. In such circumstances, active choosing increases the likelihood that people will end up with their preferred outcomes.

Active choosing might also be preferred when public officials lack relevant information, so that the chosen default rule might be harmful. This is an especially important point. If officials are inadequately informed, and if the default rule is no better than a guess, that rule might lead people in the wrong direction. The same point argues against a default rule when self-interested private groups have managed to call for it, even though it is not in the interest of those on whom it is imposed. Active choosing is much less risky on these counts.

As compared with either opt-in or opt-out, active choosing can have significant advantages when the relevant group has a great deal of diversity, so that a single approach is unlikely to fit variable circumstances. In such contexts, a default rule may also be harmful, because the power of inertia, or the force of suggestion, may mean that many people will end up in a situation that is not in their interest. For this reason, active choosing may be better.

On the other hand, active choosing can have significant disadvantages. One disadvantage is that in situations of unfamiliarity or great complexity, in which people lack information or experience, active choosing may impose unjustified or excessive burdens. These burdens include the resources required to enforce the requirement to choose and the time required for people to obtain relevant information and to make the choice. As compared with a default rule, active choosing increases the costs of decisions, possibly significantly; it also might increase errors, possibly significantly, if the area is unfamiliar and confusing. In such situations, opt-in or opt-out might produce better outcomes for people.

In the private sector, default rules are often in people's interests, and active choosing would impose unnecessary burdens. When public officials have good reason for confidence that a particular default rule will fit with the informed preferences of the relevant group, and thus promote its interests, it may be preferable to select that default rule rather than to require active choosing (Sunstein 2010a). Personalized default rules, by virtue of their accuracy, may have particular virtues on this count.

#### 4.4 Structuring Choices

Complexity can also create problems through a phenomenon known as *choice overload*. In the traditional view, having more choices helps, and never harms, consumers or program participants. This view is based on the reasonable judgment that, if an additional option is not better than existing options, people will simply not choose it. In general, more choices are indeed desirable, but an increasing body of research offers certain potential qualifications, especially in unusually complex situations (Sethi-Iyengar, Huberman, and Jiang 2004). For example, there is some evidence that enrollment may decline (Sethi-Iyengar, Huberman, and Jiang 2004), and

asset allocations may worsen (Iyengar and Kamenica 2010) as the menu of investment options in a 401(k) plan expands.

Responding to this general problem in the context of prescription drug plans, CMS has taken steps to maintain freedom of choice while also reducing unhelpful and unnecessary complexity (Gruber and Abaluck 2011). The CMS Medicare Part D program rules require sponsors to ensure that when they provide multiple plan offerings, those offerings have meaningful differences. The rules also eliminate plans with persistently low enrollments, on the ground that those plans increase the complexity of choices without adding value (see also Korobkin 2014).

## 5 SALIENCE AS NUDGE

---

It is often possible to promote regulatory goals by making certain features of a product or a situation more *salient* to consumers. As a simple example of salience effects, consider alcohol taxes. There is evidence that when such taxes are specifically identified in the posted price, increases in such taxes have a larger negative effect on alcohol consumption than when they are applied at the register (Chetty, Looney, and Kroft 2009; Finkelstein 2009). Incentives matter, but in order to matter, they must be salient. Sensible regulatory policies, especially those that involve disclosure, are attentive to the importance of salience. And if we are attentive to salience, we will be alert to the possibility that what was salient at one time might become background noise at another. Information disclosure and warnings should be attentive to this risk.

People's attention is limited, and regulatory goals are not always served merely by altering policy or disclosing information. The relevant policy or information must also be salient. In the context of fiscal policy, consider the question whether to provide payments in the form of a one-time check or instead in the form of reduced withholding. Would one or another approach lead to increased spending?

In the abstract, it may be predicted that there would be no difference as a result of delivery method. But some evidence suggests that a one-time stimulus payment may have greater effects in increasing spending than does an economically equivalent reduction in withholding (Sahm, Shapiro, and Slemrod 2011). A potential explanation, with support in the evidence, involves the importance of salience or visibility. Indeed, a majority of households did not notice the withholding changes in the relevant study, and households who found "a small but repeated boost to their paychecks" appear to be less likely to use the money for his purchases. More research is needed here.

There are many potential applications. Well-designed labels make relevant factors salient to those who will see them. The significant consequences of easy accessibility and convenience (return to the issue of obesity) can be seen as a close cousin of salience effects.

A similar point applies in the domain of energy efficiency. For many consumers, the potential savings of energy-efficient products may not be salient at the time of purchase, even if those savings are significant. The “Energy Paradox” refers to the fact that some consumers do not purchase energy-efficient products even when it is clearly in their economic interest to do so. Empirical work suggests that nonprice interventions, by making the effects of energy use more salient, can alter decisions and significantly reduce electricity use. There is evidence that such interventions can lead to private as well as public savings (Howarth, Haddad, and Paton 2000).

A related approach attempts to identify and consider the frame through which people interpret information. Research suggests that some consumers may not seriously consider annuities in retirement to insure against longevity risk—the risk that they will outlive their assets—because they do not fully appreciate the potential advantages of annuities (Brown 2007). One hypothesis is that some people evaluate annuities in an *investment frame* that focuses narrowly on risk and return (Brown et al. 2008). Looking through such a frame, consumers focus on the risk that they could die soon after annuity purchase and lose all of their money. Some evidence suggests that efforts to shift consumers into a *consumption frame*, which focuses on the end result of what they can consume over time, help consumers appreciate the potential benefits of annuities (Brown et al. 2008). The goal here is not to suggest a view on any particular approach to retirement; it is merely to emphasize that the relevant frame can increase salience.

## 6 SOCIAL NORMS AS NUDGES

---

Social scientists have emphasized the importance of social practices and norms, which have a significant influence on individual decisions. If people learn that they are using more energy than similarly situated others, their energy use may decline—saving money while also reducing pollution. The same point applies to health-related behavior. It has long been understood that people are more likely to engage in healthy behavior if they live or work with others who so engage. And if people are in a social network with other people who are obese, they are significantly more likely to become obese themselves. The behavior of relevant others can provide valuable information about sensible or appropriate courses of action. As noted above, informational cascades are a possible consequence, as people rely on, and thus amplify, the informational signals produced by the actions of their predecessors. Similarly, those actions can provide information about what others will approve and disapprove (see generally chapter 10 by Bilz and Nadler in this volume).

These points have implications for regulatory policy. For example, smoking and seat belt regulations appear to have worked hand in hand with emerging social norms, helping to reduce deaths and injuries. In the context of seat belt usage, there has been a dramatic change in behavior, with an increase in a few decades from usage rates under 15% to usage rates over 70%, in significant part as a result of social norms that

operated in concert with regulatory changes (see Sunstein 2011). In some domains, social norms have helped to promote compliance with law even without active enforcement. Public-private partnerships can be especially important in this domain, as those in the private sector emphasize norms that increase compliance with law and promote safer choices.

Consider as well the problem of distracted driving. On October 1, 2009, President Obama issued an executive order that bans federal employees from texting while driving. Such steps can help promote a social norm against texting while driving, thus reducing risks. This same approach—emphasizing social norms—might be applied in many domains. In the domain of childhood obesity, for example, a social norm in favor of healthy eating and proper exercise could produce significant health benefits. Here, as elsewhere, public-private partnerships can play a key role, with those in the private sector helping to spur emerging norms that promote better choices by and for children.

In particular, the “Let’s Move” initiative has emphasized such partnerships. First Lady Michelle Obama has collaborated with Walmart to promote healthier choices (Mulligan 2011). As part of that initiative, Walmart has committed to reformulating thousands of everyday packaged food items by 2015 by reducing sodium 25% and added sugars 10%, and by removing all remaining industrial produced trans fats. It has also committed to reduce the costs of healthier options, thus making those costs comparable to the costs of less healthy choices, and at the same time to reduce the costs of fruits and vegetables. Finally, Walmart has agreed to develop a “healthy seal” to help consumers to identify healthy choices.

In a similar vein, a number of large food and beverage companies have pledged to remove 1.5 trillion calories from their products by 2015, in an effort to combat childhood obesity (*USA Today* 2010). The relevant steps include reduction of product sizes and introduction of lower calorie foods. Finally, the Food Marketing Institute and the Grocery Manufacturers Association have agreed to promote informed choice through a “Nutrition Keys” label, designed in part to combat childhood obesity (Food Marketing Institute 2011).

---

## 7 BEYOND INCENTIVES

---

The goal of this chapter has been to outline some of the key findings in recent behavioral research and to sketch the implications for regulatory policy. A general conclusion is that while material incentives (including price and anticipated health effects) greatly matter, outcomes are independently influenced by choice architecture, including (1) the social environment and (2) prevailing social norms. When some people, cities, and nations do well and others less so, it is often because the former, and not the latter, are able to benefit from aspects of the environment, and from prevailing norms, that enable them to take for granted, and perhaps not even to think much about, a set of practices that serve them well.

While disclosure of information is an important regulatory tool, steps must be taken to ensure that disclosure will be not merely technically accurate but also meaningful and helpful. Such steps require careful attention to how people process and use information. It is useful to distinguish between summary disclosure, typically provided at the point of purchase, and full disclosure, typically provided on the Internet. Summary disclosure should be clear, simple, and salient, and it should emphasize factors that matter to people (such as annual monetary value of fuel-efficient or energy-efficient choices).

Full disclosure should provide information that can be used in multiple ways, thus improving the operation of markets; often the most important uses come from the private sector. In all cases, disclosure is most useful if it informs people of what, precisely, they might do in order to avoid significant risks or obtain significant benefits.

Default rules can greatly affect social outcomes, and in some circumstances, sensible defaults can serve as a complement or alternative to mandates and bans. One of the advantages of well-chosen default rules is that they can simplify and ease choices—for example, by producing automatic enrollment in programs that are generally beneficial while also allowing people to opt out. A potential problem is that regulators may not know which default rule is best and one size may not fit all. When the relevant group is diverse and the domain is familiar, active choosing is likely to be preferable to default rules.

Because complexity can often have undesirable or unintended side effects—including high costs, noncompliance with law, and reduced participation in useful programs—simplification may well help to promote regulatory goals. Indeed, simplification can often have surprisingly large effects. Reduced paperwork and form-filing burdens (as, for example, through fewer questions, use of skip patterns, electronic filing, and prepopulation) can produce significant benefits. It may also be desirable to take steps to ease participation in both private and public programs by increasing convenience and by giving people clearer signals about what, exactly, they are required to do.

As behavioral research has shown, people are far more likely to respond when certain facts, risks, or possibilities are salient; effective warnings take account of this fact. Finally, regulation can work in concert with social norms, helping to promote agreed-upon public goals and to increase compliance with legal requirements. Public-private partnerships, enlisting the creativity of the private sector, can be especially helpful in this regard.

## ACKNOWLEDGMENT

---

This chapter draws heavily on Sunstein (2011), and readers interested in relevant details might consult that discussion.

## REFERENCES

- Affordable Care Act. 2010. The Patient Protection and Affordable Care Act of 2010. Pub L No 111-148, 124 Stat 119, codified in various sections of Title 42.
- Allcott, H. 2011. Consumers' perceptions and misperceptions of energy costs. *American Economic Review* 101:98–104.
- Beshears, J., J. Choi, D. Laibson, and B. Madrian. 2010. The limitations of defaults. Retrieved from <http://www.nber.org/programs/ag/rrc/NB10-02,%20Beshears,%20Choi,%20Laibson,%20Madrian.pdf>.
- Bisin, A., A. Moro, and G. Topa. 2011, June. The empirical content of models with multiple equilibria in economies with social interactions. Working Paper No. 17196. Retrieved from <http://www.newyorkfed.org/research/economists/topa/multimay11b.pdf>.
- Bollinger, B., P. Leslie, and A. Sorenson. 2010, January. Calorie labeling in chain restaurants. NBER Working Paper No. 15648. Retrieved from <http://www.nber.org/papers/w15648>.
- Bronchetti, E. T., T. S. Dee, D. B. Huffman, and E. Magenheim. 2011. When a nudge isn't enough: Defaults and saving among low-income tax filers. NBER Working Paper No. 16887. Retrieved from <http://www.nber.org/papers/w16887>.
- Brown, J. R. 2007. Rational and behavioral perspectives on the role of annuities in retirement planning. NBER Working Paper No. 13537. Retrieved from <http://www.nber.org/papers/w13537>.
- Brown, J. R., J. R. Kling, S. Mullainathan, and M. V. Wrobel. 2008. Why don't people insure late-life consumption? A framing explanation of the under-annuitization puzzle. *American Economic Review* 98:304–9.
- Cabinet Office. N.d. The behavioural insights team. Retrieved from <http://www.cabinetoffice.gov.uk/behavioural-insights-team>.
- Camerer, C. F. 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
- Card D., and L. Giuliano. 2011. Peer effects and multiple equilibria in the risky behavior of friends. NBER Working Paper No. 17088. Retrieved from <http://www.nber.org/papers/w17088.pdf>.
- Carroll, G. D., J. J. Choi, D. Laibson, B. C. Madrian, and A. Metrick. 2009. Optimal Defaults and Active Decisions. *Quarterly Journal of Economics* 124:1639–74.
- Centers for Medicare and Medicaid Services. 2010, February 4. Re: Express lane eligibility option. Retrieved from <http://peerta.acf.hhs.gov/uploadedFiles/Express%20Lane%20Eligibility%20SHO%20ofinal%202-4-10%20508%20ready.pdf>.
- Chetty, R., J. Friedman, S. Leth-Petersen, T. Nielsen, and T. Olsen. 2012. Active vs. Passive Decisions and Crowdout in Retirement Savings Accounts: Evidence from Denmark. NBER Working Paper No. 18565. Retrieved from <http://www.nber.org/papers/w18565>.
- Chetty, R., A. Looney, and K. Kroft. 2009. Salience and Taxation: Theory and Evidence. *American Economic Review* 99:1145–77.
- Chiteji, N., and L. Walker. 2009. Strategies to Increase the Retirement Savings of African American Households. In W. G. Gale et al. (eds.), *Automatic* (231–60). Harrisburg, VA: R. R. Donnelley.
- CLASS Act. 2010. Pub L No 111-148, title 8, 124 Stat 828, codified at 42 USC § 300.
- Credit CARD Act. 2009. Pub L No 111-24, 123 Stat 1734, codified in various sections of Titles 15 and 16.



- Currie, J., S. DellaVigna, E. Moretti, and V. Pathania. 2010. The Effect of Fast Food Restaurants on Obesity and Weight Gain. *American Economic Journal: Economic Policy* 2(3):32–63.
- Dayan, E., and M. Bar-Hillel. 2011. Nudge to nobesity II: Menu positions influence food orders. *Judgment and Decision Making* 6(4):333–42.
- Department of Agriculture. 2011. Direct certification and certification of homeless, migrant and runaway children for free school meals. *Federal Register* 76:22785–02, 22793.
- Department of Education. 2010a. Program Integrity Issues, 75 Federal Register 66832, codified in various sections of Title 34 of the CFR.
- Department of Education. 2010b. Department of education establishes new student aid rules to protect borrowers and taxpayers. Retrieved from <http://www.ed.gov/news/press-releases/department-education-establishes-new-student-aid-rules-protect-borrowers-and-tax>.
- Department of Labor. 2011. Workplace injury, illness and fatality statistics. Retrieved from <http://www.osha.gov/oshstats/work.html>.
- Department of the Treasury. 2009, December 4. FinancialStability.gov TARP Transactions data: asset guarantee program. Retrieved from <http://www.data.gov/raw/1260>.
- Dinner, I. M., E. J. Johnson, D. G. Goldstein, and K. Liu. 2009. Partitioning default effects: Why people choose not to choose. Retrieved from <http://ssrn.com/abstract=1352488>.
- DG SANCO. 2010. Consumer affairs. Retrieved from <http://ec.europa.eu/consumers/docs/idg-sanco-brochure-consumer-behaviour-final.pdf>.
- Dodd-Frank Act. 2010. § 1021, 12 USC § 5511 (2010).
- Duflo, E., and E. Saez. 2003. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *Quarterly Journal of Economics* 118:815–42.
- Emergency Planning and Community Right to Know Act. 1986. Pub L No 99-499, 100 Stat 1728, codified at 42 USC § 11001 et seq.
- Environmental Protection Agency. 2009a. Fuel economy labeling of motor vehicles: Revisions to improve calculation of fuel economy estimates. *Federal Register* 74:61537–01, 61542, 61550–53 (amending 40 CFR Parts 86, 600).
- Environmental Protection Agency. 2009b. Mandatory reporting of greenhouse gases. *Federal Register* 74:56269–01, codified in various sections of Title 40 of the CFR.
- European Commission. 2012. Science for environment policy, future brief: Green behavior. Retrieved from <http://ec.europa.eu/environment/integration/research/newsalert/pdf/FB4.pdf>.
- Federal Trade Commission. 2009, January. Negative options: A report by the staff of the FTC's division of enforcement. Retrieved from <http://www.ftc.gov/os/2009/02/P064202negativeoptionreport.pdf>.
- Finkelstein, A. 2009. E-ZTAX: Tax salience and tax rates. *Quarterly Journal of Economics*, 124:969–1010.
- Food Market Institute. 2011. Press release: Food and beverage industry launches nutrition keys front-of-pack nutrition labeling initiative to inform consumers and combat obesity: Nutrition icon to be supported by \$50 million industry-funded consumer education campaign. Retrieved from [http://www.fmi.org/news\\_releases/index.cfm?fuseaction=mediatextandid=1207](http://www.fmi.org/news_releases/index.cfm?fuseaction=mediatextandid=1207).
- Gale, W., J. Iwry, and S. Walters. 2009. Retirement savings for middle- and lower-income households: The Pension Protection Act of 2006 and the unfinished agenda. In W. G. Gale, et al. (eds.), *Automatic* (11–27). Harrisburg, VA: R. R. Donnelley.

- Gonzales, M. H., E. Aronson, and M. A. Costanzo. 1988. Using social cognition and persuasion to promote energy conservation: A quasi-experiment. *Journal of Applied Social Psychology* 18:1049–66.
- Greenstone, M. 2009. Toward a culture of persistent regulatory experimentation and evaluation. In D. Moss and J. Cisternino (eds.), *New perspectives on regulation* (111–25). Cambridge, MA: Tobin Project.
- Gruber, J., and J. T. Abaluck. 2011. Choice inconsistencies among the elderly: Evidence from plan choice in the Medicare Part D Program. *American Economic Review* 101:1180–210.
- Hamilton, J. T. 2005. *Regulation through revelation: The origin, politics, and impacts of the Toxics Release Inventory Program*. Cambridge: Cambridge University Press.
- Healthy, Hunger-Free Kids Act. 2012. Pub L No 111-296, 124 Stat 3183.
- Heath, C., and D. Heath. 2010. *Switch: How to change things when change is hard*. New York: Broadway.
- Hirshleifer, D. 1995. The Blind leading the blind: Social influence, fads, and informational cascades. In M. Tommasi and K. Ierulli (eds.), *The new economics of human behavior* (188–215). Cambridge: Cambridge University Press.
- Howarth, R. B., B. M. Haddad, and B. Paton. 2000. The economics of energy efficiency: Insights from voluntary participation programs. *Energy Policy* 28:477–86.
- Internal Revenue Service, 2009, September. Retirement and savings initiatives: Helping Americans save for the future. Retrieved from [http://www.irs.gov/pub/irs-tege/rne\\_se0909.pdf](http://www.irs.gov/pub/irs-tege/rne_se0909.pdf).
- iNudgeYou.com. N.d. Resources. Retrieved from <http://www.inudgeyou.com/resources>.
- Iyengar, S., and E. Kamenica. 2010. Choice proliferation, simplicity seeking, and asset allocation. *Journal of Public Economics* 94:530–39.
- Jolls, C. 1998. Behavioral economics analysis of redistributive legal rules. *Vanderbilt Law Review* 51:1653–77.
- Kahneman, D., and S. Frederick. 2002. Representativeness revisited: Attribute Substitution in intuitive judgment. In T. Gilovich, D. Griffin, and D. Kahneman (eds.), *Heuristics and biases* (49–81). Cambridge: Cambridge University Press.
- Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47:263–92.
- Korobkin, R. 2014. Relative value health insurance: The behavioral law and economics solution to the health care cost crisis. *Michigan Law Review* 112:523–74.
- Leventhal, H., R. Singer, and S. Jones. 1965. Effects of fear and specificity of recommendation upon attitudes and behavior. *Journal of Personality and Social Psychology* 2:20–29.
- Levin, I. P., S. L. Schneider, and G. J. Gaeth. 1998. All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes* 76:149–88.
- McFerran, B., D. W. Dahl, G. J. Fitzsimons, and A. C. Morales. 2011. How the body type of others impacts our food consumption. In R. Batra, P. Keller, and V. Strecher (eds.), *Leveraging consumer psychology for effective health communications* (151–70). Armonk, NY: M.E. Sharpe.
- Mulligan, J. 2011, January 26. First Lady Michelle Obama announces collaboration with Walmart in support of Let's Move campaign. Retrieved from <http://www.letsmove.gov/blog/2011/01/25/first-lady-michelle-obama-announces-collaboration-walmart-support-lets-move-campaign>.

- Nickerson, D. W., and T. Rogers. 2010. Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. *Psychological Science* 21:194–99.
- Obama, B. 2009, September 5. Weekly Address.
- Office of Information and Regulatory Affairs. 2009. Information collection budget of the United States government. Retrieved from [http://www.whitehouse.gov/sites/default/files/omb/assets/inforeg/icb\\_2009.pdf](http://www.whitehouse.gov/sites/default/files/omb/assets/inforeg/icb_2009.pdf).
- Organisation for Economic Cooperation and Development. 2010. Consumer Policy Toolkit. Retrieved from <http://www.oecd.org/sti/consumerpolicy/consumerpolicytoolkit.htm>.
- Orszag, P. R. 2010, March 29. OMB, Director, SAVEings. Retrieved from <http://www.whitehouse.gov/omb/blog/10/03/29/SAVEings/>.
- Orszag, P. R., and E. Rodriguez. 2009. Retirement security for Latinos: Bolstering coverage, savings, and adequacy. In W. G. Gale et al. (eds.), *Automatic* (173–98). Harrisburg, VA: R. R. Donnelley.
- Papke, L. E., L. Walker, and M. Dworsky. 2009. Retirement savings for women: Progress to date and policies for tomorrow. In G. Gale, et al. (eds.), *Automatic* (199–230). Harrisburg, VA: R. R. Donnelley.
- Pension Protection Act. 2006. Pub L No 109-280, 120 Stat 780, codified in various sections of Titles 26 and 29.
- Redelmeier, D. A., P. Rozin, and D. Kahneman. 1993. Understanding patients' decisions: Cognitive and emotional perspectives. *Journal of the American Medical Association* 270:72–76.
- Riis, J., and R. Ratner. 2010. Simplified nutrition guidelines to fight obesity. In R. Batra, Punam A. Keller, and V. J. Strecher (eds.), *Leveraging consumer psychology for effective health communications: The obesity challenge*. Armonk, NY: M.E. Sharpe.
- Rottenstreich, Y., and C. K. Hsee. 2001. Money, kisses, and electric shocks: On the affective psychology of risk. *Psychological Science* 12:185–90.
- Rozin, P., S. Scott, M. Dingley, J. K. Urbanek, H. Jiang, and M. Kaltenbach. 2011. Nudge to nobesity I: Minor changes in accessibility decrease food intake. *Judgment and Decision Making* 6:323–32.
- Sahm, C. R., M. D. Shapiro, and J. Slemrod. 2011. Check in the mail or more in the paycheck: Does the effectiveness of fiscal stimulus depend on how it is delivered? Finance and Economics Discussion Series 2010-40. Retrieved from <http://www.federalreserve.gov/pubs/feds/2010/201040/201040pap.pdf>.
- Schwartz, J., J. Riis, B. Elbel, and D. Arieli. 2011. Inviting consumers to downsize fast-food portions significantly reduces calorie consumption. *Food and Health* 31(2):1–9.
- Sethi-Iyengar, S., G. Huberman, and W. Jiang. 2004. How much choice is too much? Contributions to 401(k) retirement plans. In O. S. Mitchell and S. P. Utkus (eds.), *Pension design and structure: New lessons from behavioral finance*. Oxford: Oxford University Press.
- Sharot, T. 2011. *The optimism bias: A tour of the irrationally positive brain*. New York: Knopf.
- Slovic, P. 1998. Do adolescent smokers know the risks? *Duke Law Journal* 47:1133–41.
- Stango, V., and J. Zinman. 2011. Limited and varying consumer attention: Evidence from shocks to the salience of bank overdraft fees. Federal Reserve Bank of Philadelphia Working Paper No. 11-17. Retrieved from [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1817916](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1817916).
- Sunstein, C. R. 2010a. Administrator, OIRA, Memorandum for the Heads of Executive Departments and Agencies, Disclosure and Simplification as Regulatory Tools. Retrieved from [http://www.whitehouse.gov/sites/default/files/omb/assets/inforeg/disclosure\\_principles.pdf](http://www.whitehouse.gov/sites/default/files/omb/assets/inforeg/disclosure_principles.pdf).

- Sunstein, C. R. 2011, February. Memorandum for chief information officers: Minimizing paperwork and reporting burdens. Retrieved from [http://www.whitehouse.gov/sites/default/files/omb/inforeg/icb/2011\\_ICB\\_Data\\_Call.pdf](http://www.whitehouse.gov/sites/default/files/omb/inforeg/icb/2011_ICB_Data_Call.pdf).
- Sunstein, C. R. 2011. Empirically informed regulation. *University of Chicago Law Review* 78:1349–429.
- Sunstein, C. R. 2013. *Simpler*. New York: Simon and Schuster.
- Thaler, R. H., and C. R. Sunstein. 2008. *Nudge*. New Haven: Yale University Press.
- Tversky, A., and D. Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5:207–32.
- USA Today. 2010, May 21. Food giants pledge to cut 1.5 trillion calories out of products. Retrieved from [http://www.usatoday.com/money/industries/food/2010-05-17-cutting-calories\\_N.htm](http://www.usatoday.com/money/industries/food/2010-05-17-cutting-calories_N.htm).
- Wansink, B. 2010. *Mindless Eating*. New York: Bantam Dell.
- Wansink, B., D. Just, and J. McKendry. 2010. Lunch line redesign. *New York Times*, A35.
- Weinstein, N. 1987. Unrealistic optimism about susceptibility to health problems: Conclusions from a community-wide sample. *Journal of Behavioral Medicine* 10: 481–500.
- White House Task Force on Childhood Obesity. 2010, May. Report to the president: Solving the problem of childhood obesity within a generation. Retrieved from [http://www.letsmove.gov/pdf/TaskForce\\_on\\_Childhood\\_Obesity\\_May2010\\_FullReport.pdf](http://www.letsmove.gov/pdf/TaskForce_on_Childhood_Obesity_May2010_FullReport.pdf).
- Zamir, E. 1998. The efficiency of paternalism. *Virginia Law Review* 84:229–86.
- Zwane, A. P., et al. 2011. Being surveyed can change later behavior and related parameter estimates. *Proceedings of the National Academy of Sciences* 108:1821–26.
- 9 CFR § 317.309.
- 16 CFR § 425
- 29 CFR §§ 1910, 1915, 1926.
- 29 CFR § 1926
- 29 CFR § 2550.404a-5 (2011).
- 34 CFR § 668.6 (2010).

## CHAPTER 29

---

# ENVIRONMENTAL LAW

---

ADRIAN KUENZLER AND DOUGLAS A. KYSAR

### 1 INTRODUCTION

---

RECENT years have witnessed the beginning of an increasingly important trend in the study and practice of environmental law and policy, namely the endeavor to integrate insights from the behavioral sciences into regulatory design and implementation. Key reports issued by academics and government agencies have discussed and recommended the use of low-cost, choice-preserving approaches to regulatory problems and proposed several ways in which such tools as disclosure requirements, default rules, and regulatory simplification can help guide and fulfill environmental policy (for a helpful overview see Sunstein 2011; chapter 28 by Sunstein in this volume). Proponents of this behavioral approach have utilized findings from cognitive psychology, behavioral economics, and other social sciences to present a comprehensive critique of conventional regulatory approaches. Alongside this critique, they have advanced their own agenda for regulatory reform, one that again is informed by behavioral insights. This agenda largely calls for the entrustment of environmental policymaking to experts in government agencies who, for their part, rely on scientific risk assessment and economic cost-benefit analysis (CBA) to identify desirable environmental policy outcomes, while using behavioralist insights to influence the design of policy tools that noncoercively guide individuals to achieve those outcomes.

This chapter offers an overview and assessment of the behavioralist turn in environmental law and policy. We begin in section 2 by putting the emergence of the behavioral approach in the historical context of developments in the theory of law and economics as applied to environmental, health, and safety regulation. From the vantage point of law and economics, the many quirks of human perception, judgment, and decision-making discovered by behavioral scientists seem to raise fundamental questions with respect to the accuracy and validity of the rational actor model at the heart of law and economics. These same cognitive phenomena seem equally problematic for policymakers, given that the individual risk perceptions lying behind societal demands

for environmental, health, and safety regulation may be of questionable empirical and normative significance. The proposed behavioralist solution to these complications has been that law and economics practitioners should adjust the rational actor model to incorporate systematic cognitive biases and other durable findings from the social sciences, and that government regulators should direct their policies toward correcting cognitive errors in public perceptions of risk through minimalist and noncoercive interventions.

In section 3, we lay out three critiques of the behavioralist approach that stay largely *within* the approach's theoretical framework. First, traditional precautionary environmental regulation—criticized by behavioralists as reflective of lay cognitive errors in risk perception—has been defended by some scholars as providing important bias correctives of its own. Second, some researchers have questioned the behavioralist distrust of lay risk perceptions, finding in such perceptions a variety of contextual factors that may be of normative importance. This diversity and indeterminacy within the behavioralist literature gives rise to a related concern regarding the manner in which the literature may be used within highly contested regulatory proceedings. Third, the behavioralist literature offers powerful reasons to believe that public demand for protection against harmful market externalities may systematically *understate* the actual societal need for protection. Thus, an alternative interpretation of the relevance of behavioral research to law undermines the supposition, shared by many contemporary economists and lawyers, that policymakers should adhere to the conventional welfare economic framework for justifying and establishing state interventions, in which a strong presumption is placed against government regulation and in favor of satisfying human preferences simply as they are given. In other words, behavioralist findings may justify not only newfangled policy “nudges,” but also traditional regulatory “shoves.”

Section 4 of this chapter uses lessons from the behavioral sciences to raise more fundamental questions about the economic approach to environmental law, whether of the traditional or the behavioralist variety. When we are forced to decide how much ecosystem structure can be converted to economic production and how much must be conserved to provide essential ecosystem services in the future, we cannot adequately do so by estimating the marginal value of environmental costs and benefits, as conventional valuation tools require. Yet this inadequacy is not just an upshot of the cost-benefit approach to environmental regulation; it also is a result of the underspecification of human nature by both conventional and behavioral law and economics. Behavioralism teaches us that environmental regulatory policy cannot be neutral in relation to human values and choices. Instead, our values and choices may be endogenously affected by the very procedures and institutions that are supposed to merely identify and implement them. When viewed “all the way down,” behavioralism tells us that, to a nontrivial extent, human beings are what society asks and expects them to be. If it asks little, it gets little. If it is more demanding, it gets more. Under those circumstances, environmental regulators should seek to (re)arrange social rules and incentives in a manner that induces people to act in ways that serve the intergenerational public good of environmental sustainability. The true wisdom of seeking that

common good must be glimpsed from outside the law and economics framework—again, whether of the conventional or the behavioral variety. Section 5 concludes.

## 2 THE BEHAVIORALIST TURN IN ENVIRONMENTAL LAW AND ECONOMICS

---

### 2.1 Early Rationalist Critiques of Environmental, Health, and Safety Regulation

From the late 1960s onwards, a number of federal statutes in the United States mandated ambitious environmental, health, and safety goals, on the principle that administrative agencies should act in a precautionary manner and promulgate the most stringent standards feasible or require the best technology achievable to protect people and the environment from harm (Lazarus 2004). A number of these early statutes turned out to be highly successful in a variety of different contexts, encouraging dramatic improvements in air and water quality, for instance. Yet many of the mandates were deliberately insensitive to the costs of achieving their goals or else included costs as a decision-making factor only as a way of establishing an outside limit to what level of protection the government could demand of regulated actors (McGarity 2004). That being the case, sustained criticism of the traditional approach arose, often through national performance reviews that gave prominence to the economic assessment of regulatory costs and benefits and the related issue of how priorities should be set within and across governmental agencies (Heinzerling 1998).

Accordingly, in the United States a shift towards CBA of federal implementation of environmental, health, and safety statutes occurred with the formal creation of a government mechanism for review of major regulations (Sunstein 2002). The use of CBA in US public administration dates at least as far back as the Flood Control Act of 1936, but CBA did not become a prerequisite for most major regulations until February 1981, when President Reagan issued his landmark Executive Order 12291 to agencies that “[r]egulatory action shall not be undertaken unless the potential benefits to society for the regulation outweigh the potential costs to society.” According to the order, each agency would have to transmit a Regulatory Impact Analysis to the Office of Management and Budget (OMB) at least sixty days prior to the finalization of a major rule. If the OMB Director deemed an agency’s CBA to be insufficient, the Director could “[r]equire an agency to obtain and evaluate. . . additional relevant data from any appropriate source” (Executive Order 12291). The basic thrust of the order has survived three changes in partisan control of the White House and today the CBA approach substantially influences environmental regulation even in contexts where the underlying statute being implemented seems to reject such economic balancing (Kysar 2009).



Beyond these developments in the United States, regulatory impact analysis has become a widespread global phenomenon. From the 1990s onwards, international bodies such as the OECD, the WTO, and the European Commission began to rely on empirical techniques of decision-making, and most of them have explicitly mandated or encouraged regulatory impact analysis programs—with CBA at their core—for use in policymaking (Jacob et al. 2011; OECD 1997). Today, regulatory impact assessment constitutes an essential part of governance in many countries, though the scope and effect of cost-benefit assessments varies across different regimes (Livermore and Revesz 2013).

Behind the push toward CBA lies a competitive market vision in which mandating a single compliance technique or quality standard, as early federal statutes in the United States had done, fails to appreciate the diversity and innovative capacity of market actors, thereby precluding opportunities for companies to achieve the same level of protection through other, less costly means. Likewise, uniform environmental or health quality standards are thought to be insensitive to the overall social welfare gains that might follow from variation in standards based on factors such as population density, individual preference, budget constraints, and so on. More generally, traditional regulatory approaches are seen to require ever-greater levels of protection up to the point that compliance costs begin to sharply rise, even though more protection might not always be better from the perspective of aggregate social welfare (Revesz and Livermore 2008). Indeed, the central insight of the economic approach to environmental law and policy is that the “right” level of environmental benefit to pursue can only be determined by evaluating the costs of attaining it.

To be implemented, this abstract model of CBA requires a method for converting the protected subjects of environmental, health, and safety law into monetary equivalent values in order to enable quantitative comparison with regulatory costs. Such commensuration is understood to occur at least implicitly any time we draw a regulatory line at something less than absolute protection. Moreover, because any kind of protective regulation can itself have unintended, risk-enhancing consequences, or because the product or activity under regulatory scrutiny may provide potential environmental benefits in addition to its suspected risks, analysts contend that most regulations involve risks on either side of the regulatory equation, whether or not nature is “commodified” explicitly during the regulatory design process (Graham and Wiener 1995; Viscusi 1994). Thus, making the advantages and disadvantages of different courses of action explicit and commensurable through CBA is believed to render environmental policy more rational, transparent, and realistic as compared to the lofty aspirational standard of early federal statutes.

To evaluate the benefits and costs of regulatory proposals, regulators could utilize one of several different forms of welfarism, such as those that rely on hedonic or objective list accounts of human well-being (Kagan 1998). In most real-world applications, however, CBA is implemented as a form of preferentialism in which well-being is equated with the satisfaction of individuals’ preferences. Thus, the economic approach generally relies on estimates of the amount individuals are willing to pay for a gain (or

to avoid a loss) or the amount they are willing to accept as compensation for a loss (or to forgo a gain). In order to infer such amounts from individual behavior, the economic approach presupposes that discrete choices revealed through market transactions are rational manifestations of individuals' preferences (Becker 1976). On that basis, individual actors are thought to be best able to choose from the options that yield the highest net benefits to themselves, subject to their own budget constraints. Utilizing the best available evidence of individual preferences when evaluating the costs and benefits of government regulation therefore rests on a strong liberal premise: Rather than presume individuals want some fixed level of protection irrespective of cost, the CBA approach aims to give them the level they prefer as revealed by their behavior in other settings.

The most prominent and practically significant manifestation of the revealed preference approach is the academic literature on the monetary value of a statistical life (VSL). When assessing environmental, health, and safety regulations, US federal agencies typically use a monetary value of the benefits of risk avoidance that has been empirically derived through observations of individual behavior in labor markets. The central theoretical premise of these studies is that, other things equal, workers demand a higher wage in order to accept a higher level of occupational fatality risk. Large data sets have been analyzed with this assumption in mind and a consensus view among practitioners has developed that the implicit value of a statistical life revealed by wage premiums lies somewhere around \$7 million (Viscusi and Aldy 2003; Kochi, Hubbell, and Kramer 2006). Such figures now are routinely used to compute the monetized benefits of proposed regulations, affording in the process a ready basis for comparison against the estimated economic costs of regulation. Referencing this literature, analysts contend that citizens should be granted only that level of environmental, health, and safety benefit that they themselves have proven willing to purchase through their revealed market behavior. Again, both practical and principled arguments support this approach to welfarist policy analysis, as alternative welfare theories (such as hedonic or objective list accounts) often lack comparably rich empirical literatures to support real-world policy implementation, and as the preferentialist approach seems to comport with the liberal ideal of respecting the autonomous decision-making of individuals (Adler and Posner 2006).

## 2.2 The Behavioralist Critique of Individual Rationality

A multitude of empirical findings about human decision-making behavior have called into question the assumptions of the neoclassical model of "homo economicus" (Jolls, Sunstein, and Thaler 1998). Studies have found that people often base their decisions on psychological heuristics, which are essentially rules of thumb used to process uncertain or information-rich options (Kahneman 2003). These shortcuts can lead to results that deviate from the predictions of standard economic or rational choice theory because people give too much weight to available or easily accessible information,

orientate themselves toward what is familiar or prominent, allow themselves to be misled by highlighted similarities, or tend to confirm previously held beliefs even if they are inaccurate. In many predictable real-world settings, people therefore fail to make optimal decisions that maximize their subjective expected utility, casting doubt upon the use of individual preferences and behavior as the measuring rod of welfare within the economic analysis of law (Korobkin 2009).

For this reason, even though CBA is typically justified as a method of promoting social welfare-maximization, behavioralists have argued that its most plausible rationale is based on cognitive grounds—as a mechanism for counteracting predictable shortcomings in individual and social cognition that would otherwise, via democratic processes, result in inefficient laws and policies. CBA, in their view, serves as a corrective to poor judgments by individuals and societies that result from the use of cognitive heuristics, from informational and reputational cascades, from thinking processes in which benefits are “on screen” but costs are not, from intense emotional reactions, and so forth (Sunstein 2000). Traditional precautionary regulatory approaches, in contrast, seem to reinforce these deficits as they appear to focus asymmetrically on the risks of a proposed product or activity, to the exclusion of benefits. Precautionary approaches thus are thought to result in policy choices that potentially cause more harm to human health or the environment than good (Sunstein 2003).

This shift in perspective away from a strict efficiency-based point of view to a cognitive characterization of CBA depends on an array of intercoupled psychological mechanisms that are believed to culminate in systematic distortions of public risk perception. Among the most notable are the availability heuristic, probability neglect, loss aversion, status quo bias, and affect. A number of reinforcing factors in particular social settings have also been identified. These include availability cascades, moral panics, and group polarization. Each of these phenomena will be examined briefly below.

When evaluating a risk, people rely upon how easily they can think of examples of the potential misfortune (Slovic 2000). Familiar events are sometimes more cognitively available than rare events simply because the former are more numerous, but relative salience or vividness also plays an important part in availability. If floods have occurred in the immediate past, people who live on flood plains are far more likely to purchase insurance, but the demand for insurance tends to decline steadily as vivid memories recede (Sunstein 2005, with reference to Tversky and Kahneman 1982 and Slovic 2000). Similarly, the effect of seeing a house burning down has a greater impact on the subjective probability assessment of such accidents than the impact of reading about a fire in a local paper (Tversky and Kahneman 1982). At the societal level, analysts worry that the demand for environmental, health, and safety protection may be driven by vivid, dramatic events that bear little resemblance to the risk priorities that would follow from a detached ranking based solely on actuarial data (Sunstein 2003).

Public perceptions of risk also may be affected by probability neglect. As the behavioral research has shown, even if highly improbable, people may gravitate towards worst-case scenarios and support costly measures to guard against them, however remote the associated risks (Sunstein 2005). Thus, often people may incur too high a

cost to avoid low probability outcomes if fear or salience is involved (Rottenstreich and Hsee 2001). The precautionary regulatory approach, in light of this finding, may produce serious difficulties for agencies, as regulators may—no less than individuals—obsess over low-probability risks in a way that can lead to costly expenditures for little or no gain. Worse still, new risks may arise as a result of the regulatory intervention, with an equal or even greater probability than those initially addressed.

In addition, the behavioral research has shown that people see a loss from the status quo as having more impact on their well-being than an equally sized gain from the status quo. As Richard Thaler and other researchers demonstrated in classic experiments involving loss aversion and the endowment effect, facing a loss can make us afraid in ways that exceed our moods of positive anticipation when we look forward to some supplement to what we now have (Thaler 1991; chapter 12 by Korobkin in this volume). With respect to fear and risk regulation, Cass Sunstein therefore suggests that “[p]eople will be closely attuned to the losses produced by any newly introduced risk, or by any aggravation of existing risks, but far less concerned with the benefits that are foregone as a result of regulation” (Sunstein 2005, p. 42). A recommendation to take precautions against potential losses from the existing state of affairs will therefore seem more appealing than an anticipation of potential benefits that would flow from changing that state of affairs. We are biased toward the status quo: When confronted with a potentially beneficial but also risky course of action, we fall back on the precautionary instinct that it is “better to be safe than sorry” (Sunstein 2005, p. 47; chapter 11 by Zamir in this volume).

Finally, perhaps the most fundamental of all the mechanisms of risk perception mentioned is affect. The emotional reactions that people experience in response to ostensibly threatening incidents have been found to be some of the most reliable indicators of how risky individuals perceive them to be (Slovic et al. 2004; Finucane et al. 2000; Loewenstein et al. 2001). What is readily available to some individuals depends in large part upon how emotionally captivating the impressions of misfortune generated are. As has been shown, many of those who reject gun control legislation are alert to incidents in which private gun ownership allowed people to fend off criminal violence (Kahan and Braman 2003). Similarly, researchers believe that people display loss aversion or status quo biases because when they anticipate losing what they now have, they can become genuinely afraid and fixate their attention on the unfavorable outcome instead of on its likelihood (Sunstein 2005). Seen in this light, the language of benefits and costs is thought to be able to “cool down” people’s intuitive, affect-driven reactions—just as more emotive discourses, such as the precautionary principle, appear to stir them up (Kysar 2011).

As mentioned at the outset of this section, proponents of the behavioral economic approach have presented a set of additional factors that magnify the distorting influences just described. The first involves social bandwagons and cascades, in which apparently representative anecdotes and gripping examples are spread rapidly from one person to another (Heath 1996). Processes of this sort may have played a role in public conversations over mad cow disease, Asian flu, and other prominent risks

(Sunstein 2005), when people observed the actions of their peers and then reached the same or similar perspectives, regardless of their own individual information signals. The second type of reinforcing factor involves group polarization. In the context of risk evaluation, group polarization occurs when like-minded people engage in deliberations over risks and how to diminish them. In the course of such debates, people often “end up accepting a more extreme version of the views with which they began” (Sunstein 2005, p. 98). They move in the direction of the most persuasive and frequently defended position in a group so that in the context of, say, climate change, group members who believe in the threat of climate change will collect a disproportionate number of arguments justifying their fear. The group as a whole will then shift toward greater fear. Moreover, because group members will want to be perceived favorably by others they will adjust their attitudes towards the dominant position. When other people share some group members’ views, they then grow more confident that they are correct and their views become increasingly extreme (Baron et al. 1996; Hatfield, Cacioppo, and Rapson 1994). Analysts believe that these catalyzing and distorting dynamics play a powerful role in the demand for environmental, health, and safety regulation. Again, the precautionary approach to regulation is criticized as serving to enable fear cascades and overreactions, rather than tamper them down through cool-headed analysis.

The preceding list of distorting influences is far from exhaustive. Nevertheless, it should make plain the ways that cognitive processes are believed to shape and distort risk perceptions. The implication is that not only are people alarmed when they ought not to be, but they also can be ignorant when they should be frightened (Sunstein 2005). A society that fears nuclear power because of salient and dreaded meltdown scenarios might well pay too little attention to the risks associated with coal- or gas-fired power plants. And societies that ban genetic modification of food because they fear the unfamiliar may be insufficiently mindful of the potential benefits of the technology (McHughen 2000). The upshot is not that it becomes possible to say whether excessive fear or excessive fearlessness dominates in public risk perception. Rather, the conclusion that has been drawn from this account is that the public, driven by emotional reactions and cognitive biases, simply cannot be expected to get it right most of the time. In fact, proponents of the economic approach see a key task of government regulation to be the uncovering of irrationality in public risk perceptions—a task that in their view is best achieved through systematic analyses of benefits and costs of regulation, carried out by scientific experts and government agencies that are relatively shielded from public demands.

The behavioralist account thus rejects a populist system of regulation that takes public risk evaluations at face value (Kahan et al. 2006). Well-functioning governments, it is claimed, aspire to be “deliberative democracies” that account for the public’s anxieties, but “their responsiveness is complemented by a commitment to deliberation, in the form of reflection and reason giving” (Sunstein 2005, p. 1). Institutions and procedures ought to be created that feature a degree of insulation in order to resist public demands that seemingly lack a solid foundation. Experts are supposed to be relatively immune from the influences that inevitably distort public evaluations of risks because

experts, by virtue of their role and training, can be expected to use more calculative and evidence-based reasoning, and thus be better equipped to assess whether risks are real and worthy of regulation (Sunstein 2005).

### 2.3 The Behavioralist as Regulator

Findings such as those just reviewed have also sparked controversy over whether the law should provide individuals with assistance when making decisions: If we really do stagger through life, being pulled and shoved by our appetites, whispered to by fears and beckoned by hopes, mostly keeping our balance but regularly obtaining less utility than is hypothetically conceivable, then might the best social response be a form of legal paternalism—that is, the position that one of the legitimate goals of law is to protect us from ourselves in particular cases, even against our own will if need be? The case for strong paternalistic intervention is subject to a number of familiar objections, the most serious being the restriction of freedom suggested by state interference with individual preferences. A regulation mandating removal of potentially toxic chemical residues in foodstuffs, for example, might harm those who would rather pay less for food and spend the saved money on something else, even if the regulation would otherwise benefit many. Legal paternalism also seems to ignore the possibility that in some circumstances it may be best for an individual to make a mistake and suffer consequences in order to learn from experience (Regan 1983). For instance, someone may, *ex ante*, value being free to choose between a hazardous workplace and a safer alternative with lower wages, but if he chooses safety then his preference for safety may increase, leaving the value he attaches to being left free to opt for risk diminished *ex post* (Burrows 1993; chapter 7 by Mitchell in this volume).

These examples highlight the problematic assumption of superior state knowledge that paternalism is based upon, and the associated postulate that a government official acting on the people's behalf is better able to maximize citizen well-being than citizens themselves. Such issues of paternalism arise in almost all contexts of regulation. Legal academics and policymakers have therefore moved toward a vision of state influence that does not resort to actual coercion and its accompanying restraints on freedom. Early proponents called such a regulatory paradigm “asymmetric paternalism” because large benefits may be created for those who are subject to cognitive errors, while imposing little or no harm on those who are not (Camerer et al. 2003). George Loewenstein and Emily Haisley set forth a comparable agenda, which they label “light paternalism” (Loewenstein and Haisley 2008). Richard Thaler and Cass Sunstein use the term “libertarian paternalism” to delineate a similar policy conception (Thaler and Sunstein 2009; Sunstein and Thaler 2003). In particular, they pursue a vision of noncoercive state intervention by setting down a “golden rule” of libertarian paternalism: The law should offer behavioral “nudges” that are most likely to help and least likely to inflict harm (Thaler and Sunstein 2009, p. 74). In the environmental regulatory context, this new solution often amounts to libertarian policy interventions that



accentuate the *social* cost of pollution, or the *social* benefits of mitigation—a policy that has also been labeled as “libertarian welfarism” (Korobkin 2009, pp. 1671–73)—as opposed to the *private* costs and benefits of such activities. Both paradigms avoid mandating particular behavior and instead merely structure the context of choice so that the negative effects of individual behavior—whether private or social—are made visible, thus making the optimal activity more desirable for individuals.

These maxims are operationalized through a choice of instruments that is best characterized by rising intensity of intervention—and thus growing restraints on freedom: (1) the use of disclosure as a regulatory tool; (2) simplifying choices through sensible default rules and reduced complexity; (3) increasing the salience of certain factors or variables to promote better decision-making; and (4) promoting desirable social norms (Sunstein 2011). These four recommendations will be discussed in turn.

The first recommendation fits comfortably within the traditional economic paradigm, as it merely seeks to overcome problems of information asymmetry or information failure that prevent actors from making welfare-maximizing decisions. In the United States, for instance, the Environmental Protection Agency (EPA) has required that car manufacturers post the estimated gas mileage and annual cost of gasoline on the windshields of new cars. Since the cost of operating a new car is presumably relevant for many buyers, and because of the trouble they would face estimating this cost on their own, such a requirement is likely to make relevant information more accessible and to enable buyers to purchase cars that are less costly to operate (Korobkin 2009). Placing a greater emphasis on enhancing overall *social* as opposed to merely *private* welfare would insist on requiring car manufacturers to add another windshield sticker, providing information regarding the car’s greenhouse gas emissions and other pollution externalities. In the same vein, a report by the Office of Information and Regulatory Affairs (OIRA) to Congress on the costs and benefits of federal regulation has suggested new initiatives along the lines of EPA’s issuance of a Greenhouse Gas Reporting rule, requiring disclosure by the most significant emitters of climate-warming pollutants (OIRA 2009). According to OIRA, the data will allow businesses to track their own emissions and to compare them to similar facilities, and will help to identify cost-effective ways to reduce emissions in the future.

The second recommendation, which more clearly reflects the behavioralist turn in law and economics, involves the insistence that regulators should pay heed to the “choice architecture” that individuals face when making many important, welfare-affecting decisions (Thaler and Sunstein 2009). For instance, when individuals encounter a significant degree of unfamiliarity and complexity for a given choice, setting a default rule by law may work effectively to “nudge” many individuals to welfare-enhancing choices. Setting a default rule takes advantage of status quo bias, as people are more prone to choose what they perceive to be part of the background or “standard” set of conditions (Kahneman, Knetsch, and Thaler 1990). Having electricity customers signed up by default to participate in a renewable energy development fund, for instance, is expected to lead to higher participation rates than a default rule in which they must opt into the program. In other contexts, regulators might simply



structure the choice architecture to require individuals actively to reach a decision, without preferencing one option over another. Such active choosing might be preferred when the government lacks relevant information so that a particular default rule might be harmful to some or many, or when the learning effects of active choosing by individuals are especially desirable (Sunstein 2011).

The third recommendation reflects the finding that the manner in which information is presented to individuals can be as significant as the decision whether to present information at all. Recognizing that actual humans are more psychologically complex than the rational information processor at the heart of the conventional economic model, proponents of behavioral law and economics insist that regulators should consider social scientific research regarding the best manner of information presentation to ensure the effectiveness of their interventions. For instance, the National Highway Transportation Safety Administration (NHTSA) in the United States sought to implement a requirement of the 2007 federal energy bill that the agency create a Tire Fuel Efficiency Consumer Information Program, consisting of both a rating system of tire fuel efficiency and a labeling program to convey the information to consumers. Although the agency had conducted some focus group and survey testing of its proposed labeling program, OIRA rejected it because the agency had not conducted “scientifically valid experiments” that were capable of quantifying “consumers’ understanding of the label” (quoted in Kysar 2011, p. 52). As this example shows, the insistence that regulation be premised upon rigorous and current social scientific understandings can itself serve to delay the adoption of regulations. Nevertheless, the ultimate goal of such insistence is the creation of consumer information programs that will be more effective than bland disclosure because they will be premised on a more sophisticated psychological understanding of how individuals perceive, process, and use information.

The fourth and final recommendation accounts for the fact that individuals are social beings whose preferences may be interdependent or may be influenced by perceived social norms (Vandenbergh 2005). For instance, household electricity consumption might go down if we nudge people with emoticons on their utility bills that reveal whether their usage is higher or lower than that of their neighbors (Ayres, Raseman, and Shih 2009). Policy measures such as these are built on a relatively straightforward mechanism: If (respected) people in a community engage in behavior that increases social welfare and that behavior is visible, a norm may emerge that increases the likelihood that others will act in accordance with the norm. Again, from the perspective of libertarian paternalism (or welfarism), such dynamics are preferable to outright regulatory command, as the possibility remains for individuals to defect from the social norm. (Indeed, one study showed that Republican electricity consumers actually *increase* their usage in response to an emoticon-based nudge program that was designed to encourage energy conservation; see Costa and Kahn 2010).

From the libertarian paternalist or welfarist perspective, a general theme of regulation emerges that is focused on small, inexpensive initiatives that have strong empirical foundations (Sunstein 2011). The justification for state intervention based on this

regulatory regime is two-tiered: first, provisional justification for interference with individual preferences requires the existence of relevant deficits in rationality; second, further endorsement of preference interference is made contingent on the possibility that the same cognitive biases can be counteracted by simply structuring the choice architecture that people are given without coercing those who do not suffer from the cognitive bias.

In sum, much of the behavioral law and economics turn within policymaking represents a confluence of, on the one hand, the neoclassical economic goal of assessing and maximizing efficiency gains as measured by preference satisfaction and, on the other hand, the behavioralists' empirical observation that individual decision-making is sometimes inconsistent with key assumptions of rational choice theory. From the regulatory theory perspective, a (more comprehensive) response was needed to the various ways in which observed behavior appears inconsistent with key assumptions of rational choice theory. Both academics and government administrators thus embarked on a common project aimed at incorporating findings from the cognitive sciences into the use of CBA and the oversight of federal regulation (Sunstein 2011), expanding its scope to important domains such as the assessment of agency responses to a wide range of environmental issues (Livermore and Revesz 2011). And although the United States has been the most prominent venue for these developments, it has been far from alone (Livermore and Revesz 2013).

### 3 BEHAVIORALIST COUNTERTURNS TO THE BEHAVIORALIST TURN

---

#### 3.1 Precautionary Regulation as a Debiasing Measure

Leading proponents of the behavioral approach to environmental regulation have demonstrated that certain psychological tendencies might lead individuals and groups to overreact to salient risks. In turn, the traditional precautionary approach to risk regulation is thought to reflect and perhaps exacerbate such tendencies by focusing asymmetrically on the harm that might come from an activity, rather than on the benefits it might pose or the harms that might follow from regulating it. This account, however, depends on a selective reading of the social scientific literature on risk perception. For instance, although much has been made of the availability bias as a catalyst for *overinvestment* in risk mitigation (Kuran and Sunstein 1999), availability errors can just as easily lead to *underinvestment* in disaster prevention. In the event of some underrated risks, the precautionary approach might allow regulators to overcome the problem of nonavailability. Consider again the case of earthquake insurance. As noted above, in the aftermath of an earthquake, demand for insurance rises, but thereafter declines steadily as recollections fade (Sunstein 2006). If governments require insurance

policies to have long-term coverage periods (or if they require a lengthy period as the default term of a contract), then individuals' efforts to reduce exposure to catastrophic risk in the immediate aftermath of a disaster might better match actuarially preferred coverage levels even when memories recede.

More generally, there are reasons to believe that human risk perception and decision-making are skewed *against* environmental protection, notwithstanding occasional availability cascades that drive concern about certain salient risks. As David Dana has noted, environmental policy decisions are often framed as a choice between "avoidance of a relatively sure, immediate, or very near-term loss of money (i.e., the costs of taking regulatory action) and avoidance of a relatively unsure, non-immediate loss in human or ecological health and welfare" (Dana 2003, p. 1320). If decision-makers are asked to juxtapose economic losses that are relatively certain with health and environmental losses that are not, and when the first are often immediate while the second relate to the future, then their judgments skew toward avoiding the former and underestimating the latter (Dana 2003). As Daniel Kahneman and Amos Tversky have shown in their seminal article on prospect theory, people generally overweigh outcomes that are considered certain relative to outcomes that are merely probable (Kahneman and Tversky 1979). In the regulatory context, future health and environmental losses are typically uncertain, both at the level of determining whether perceived threats will cause harm and with respect to the question of whether available technology or regulatory tools will in fact prevent such harm (Krier 1994; Rachlinski 2000). It is, for instance, exceedingly difficult to predict exactly what impact curbing or prohibiting deforestation will have on the magnitude and speed of climate change. By contrast, it is certain that some opportunity costs will be incurred as a result of such regulation (Dana 2003).

What is more, the behavioral research has shown that people value the avoidance of instantaneous or almost-instantaneous losses far more strongly than the avoidance of remote losses, even in the not-too-distant future (Loewenstein and Prelec 1992; Shelley 1994; Vlek and Keren 1992). As in the example above, losses in human health and associated safety hazards are typically much more distant than are job losses or forgone industrial profits. Decades can pass between exposure to a contaminant and the manifestation of a disease (Revesz 1999; Viscusi 2000). Or it may take years for us to understand the full implications of, say, the displacement of conventional plant species by genetically modified agricultural products. On the other hand, no one seems to doubt that a new regulatory burden on, for example, polluting vehicles will result in greater economic costs through increased equipment expenses, manufacturing delays, or the like (although the evidence suggests those costs are often significantly overstated by opponents of regulation; see, e.g., Morgenstern, Pizer, and Shih 2002; Berman and Bui 2001; Bezdeka, Wendling, and DiPerna 2008). The only issue that typically generates controversy is whether such measures would actually result in the prevention of health and environmental loss (Dana 2003).

To be sure, these dynamics do not tell us whether a policy based on precaution is more or less desirable than the CBA approach. The answer depends on whether the

problem of *overregulation* in the immediate aftermath of a disaster (i.e., when the relevant risk is readily available) is greater or less than the problem of *underregulation* for those risks that are not cognitively available. A large research literature in psychology tells us, however, that in the environmental regulatory context we should probably be more concerned with the latter problem. According to what are known as “dual-process theories of thinking,” for instance, *numerical* representations of saved lives often lack compelling meaning to individuals, so that statistical information will be underattended to in their decision-making and beliefs (Loewenstein et al. 2001; Slovic et al. 2002). The same psychological research suggests that we often have trouble appreciating the meaning of lost lives as their number becomes larger, experiencing a form of “psychic numbing” (Slovic et al. 2013; Fetherstonhaugh et al. 1997). In at least some risk contexts, the precautionary principle can therefore be justified pragmatically, as a method of underscoring the importance of attending to environmental issues that might otherwise be overlooked. By shifting the burden of proof toward the beneficiaries of potentially risky activities, it may work as a “debiasing” measure, enabling serious consideration of regulatory action to address otherwise underappreciated risks. The precautionary principle, in this sense, serves as a reminder that we must worry about harm to human health and the environment even—especially—if it seems distant and uncertain; it lessens biases and may result in sound policy recommendations, which otherwise tend to overemphasize the concern to avoid the certain and immediate economic costs of regulation.

### 3.2 The Indeterminacy (and Manipulability) of the Behavioral Literature

Even within the behavioral literature, researchers disagree extensively on such matters as “when lay-expert divergences represent valid expressions of public values, as opposed to cognitive errors or misperceptions” (Kysar 2003, p. 1782) and whether heuristic decision-making more generally is better seen as prone to systematic error or as remarkably well adapted to the actual environments within which humans exist (Gigerenzer, Todd, and the ABC Research Group 1999; Kysar et al. 2005). In the context of risk regulation, for instance, people seem to care not just about expected annual mortality or the aggregate amount of lives saved, but also about a range of contextual features such as the catastrophic nature of the risk in question, whether the risk is incurred voluntarily, whether it involves irreparable losses, how equitably distributed or concentrated on innocent victims the danger is, how well we understand the risk, whether it would be faced by future generations, and so forth (Sen 2000). In the eyes of some behavioral researchers, many of these components are normatively significant, such as when people are less tolerant of involuntary exposure to air pollution from a hazardous waste disposal facility than from voluntary smoking (Fischhoff et al. 1978; Heinzerling 1998).

Hence, differences in risk perception may often result from different valuations of the “same” level of risk, rather than from diverging information, factual knowledge, or mental shortcuts. Indeed, people’s general ordering of familiar threats in terms of risk is often quite accurate in the actuarial terms that experts use, such as the amount of injuries and deaths they annually cause (Morgan 1993). But a divergence occurs when people are asked to rank these hazards in terms of importance, a more open-ended frame that allows their qualitative beliefs and normative judgments to enter into the evaluation (Morgan 1993; Hill 1992). It may thus be perfectly rational and understandable when the public demands that more resources be spent to achieve the same level of risk reduction if some risks are involuntarily imposed rather than voluntarily assumed, or when risks arise through processes perceived as illegitimate rather than legitimate. People are apparently concerned about *how* they live, *how* they might die, and *who* might be in charge of their quality of life. A myopic focus on lives saved would be insensitive to issues of this kind (Boyd, Kysar, and Rachlinski 2012).

Keeping these observations in mind, some seeming puzzles in risk regulation dissolve. Even though the economic approach derives much of its appeal from being a disciplined analytic tool for evaluating regulatory choices, it runs the risk of presupposing that the world outside agencies is entirely irrational, and that choices made on any other basis than those founded on narrow risk assessment and economics must be fundamentally flawed (Gayer and Viscusi 2012). It then comes as little surprise if EPA and the US Department of Transportation are found to assume that energy efficiency is always the paramount product attribute for durable consumer goods, and that consumers’ decisions made on any other basis (such as valuations of car attributes that fuel-efficient cars do not offer) are essentially unsound (Gayer and Viscusi 2012). Read broadly, the behavioral economic approach might justify a form of government intrusion that simply overrides consumer preferences regardless of their source. But once we realize that the expert’s conceptions of value may be at odds with well-considered public perceptions of risk and benefit that are not properly dismissed as biased, then narrow technocratic assessments become much more difficult to endorse unreservedly.

In response to these differences, proponents of the economic approach have argued that experts could simply quantify the various reasons that determine laypeople’s judgment of risk and adjust cost-benefit calculations accordingly. The government’s task would then be to distinguish between lay judgments that result from simplistic or misleading thought processes (for example, as a consequence of the availability heuristic) and lay judgments that result from a different or deeper form of rationality (for example, as a result of the view that distributive equity in risk imposition matters even at the expense of aggregate social welfare). But is the economic approach actually capable of incorporating the observed disparities? To some extent, the answer is yes. The EPA, for instance, has proposed the use of a 50% “cancer premium” to value the benefits of regulations that avoid deaths by cancer, in light of empirical evidence suggesting that individuals value such avoided mortality higher than the accidental workplace risks that dominate the VSL literature. Unlike workplace accidents, many cancer deaths are

seen to be involuntarily imposed and are strongly associated with dread (Livermore and Revesz 2011).

If accepted, the “cancer premium” would work to effectively raise the permissible stringency of cancer-avoiding regulations by increasing the monetary benefits of such regulations for purposes of CBA. But the very promise of proregulatory benefits from the behavioral literature also suggests the literature’s diversity, indeterminacy, and potential manipulability: Just as behavioral researchers are unable to agree among themselves whether public demands for risk regulation are generally untrustworthy or not, interest groups within regulatory battles seem likely to cast about within the behavioral literature in search of studies and findings that support their preferred outcomes. The behavioralist turn in the law and economics approach to environmental regulation might then simply add another layer of gamesmanship to a rulemaking process that already has been aptly described as a “blood sport” (McGarity 2012).

Consider the EPA’s experience attempting to regulate coal ash waste. In this long overdue rulemaking, the EPA proposed to regulate coal ash that is a byproduct of electricity production and that contains a variety of heavy metals, carcinogens, and other materials known to be hazardous to human health. Over 100 million tons of the waste are generated each year in the United States and threaten to contaminate groundwater and cause harmful ecological and human health effects, both by leaching from unlined waste impoundments and through catastrophic releases such as the 2008 incident at the Kingston Fossil Plant in Tennessee. The EPA originally recommended that coal ash be listed as a Subtitle C “hazardous waste” and subjected to the requirements of the Resource Conservation and Recovery Act (RCRA). Following OIRA review, however, the agency revised the proposal to include both its original plan and two other significantly less stringent alternatives that would essentially treat coal ash as little different from household waste.

The primary justification for this adjustment of the RCRA requirements for coal ash came from a behavioralist argument by industry that classification of coal ash as a hazardous waste might give rise to a “stigma effect” whereby the beneficial reuse of coal ash would no longer be economically viable. At present, some coal ash is recycled and used in road paving, construction concrete, and other safe and beneficial applications. Even though the EPA was careful to note in the proposal that it was not altering an earlier determination that such beneficially reused coal ash is exempt from RCRA’s hazardous waste regulations, industry representatives argued that stringent regulation of remaining coal ash would unduly “taint” beneficially reused waste material. The price tag for this unintended behavioral consequence of regulation was \$233.5 billion in lost benefits, an amount that radically tilted the EPA’s original cost-benefit estimates against regulation.

Empirical support for the coal ash “stigma effect” is surprisingly thin, despite the significance afforded it by OIRA. One study often cited in the behavioral economic literature, for instance, found that subjects were unwilling to drink juice from a glass that had recently contained a cockroach, even though the subjects were assured that the



cockroach had been “medically sterilized” (Jolls and Sunstein 2006, citing Rozin 2001). That same study was relied upon by economists in a study of Superfund site designation to construct a model of human behavior in which “people replace calculations of risk versus benefit with a simple heuristic of shunning, the avoidance of the stigmatized object” (Schulze et al. 2004, p. 23). Just as the heuristics and biases literature more generally has been interpreted to suggest that individuals often misconstrue and mispursue their best interests, the EPA economists concluded that “while shunning may have evolved from an adaptive response to avoid contaminated food, it can be triggered in inappropriate circumstances” (Schulze et al. 2004, p. 24). Thus, because subjects were unwilling to drink juice from a glass that had contained a “medically sterilized” cockroach, individuals more generally were assumed to be prone to overreaction to risk. Consistent with this assumption, the EPA economists found that Superfund cleanup delays of ten to twenty years increase the chance of stigmatization for properties surrounding designated sites, such that the benefits of an eventual cleanup may never be reflected in restored property values.

These studies seem far afield from the context of coal ash reuse. Even the Superfund study sheds little light, as the event triggering “taint” in that context was a designation of hazardous status by the EPA and a subsequent failure to clean the site up for one or more decades. In the coal ash context, on the other hand, the EPA has specifically *exempted* beneficially reused materials from hazardous designation. Why then was industry allowed to extrapolate from a study on cockroach juice to a multibillion-dollar claim about coal ash reuse? In this rulemaking at least, the addition of behavioral sciences seems to have given industry more weaponry with which to delay regulations and fudge calculation of their impacts, rather than to improve the understanding and transparency of policy assessment.

This finding suggests that behavioral law and economics proponents must remain attentive to the institutional context within which their recommendations will be received. Even if the incorporation of behavioralist research into regulatory rulemaking is advisable from an academic perspective, the incorporation will still occur within a political setting capable of distorting the research’s implications.

### **3.3 The Demand for Regulation and the Need for Shoves, Rather Than Nudges**

Advocates of the behavioral economic approach maintain that loss aversion, the status quo bias, and the endowment effect can blind people to the benefits of potentially hazardous courses of action while keeping them closely attuned to the losses produced by a newly introduced risk or aggravation of an existing risk. Legal scholar Cass Sunstein uses the phrase “Laws of Fear” to denote the immediate and remote impacts of these phenomena on a disproportionately growing demand for regulation (Sunstein 2005). According to Sunstein, “these mechanisms show the sense in which the relevant



blindness are not arbitrary or coincidental”—they “have an unmistakable structure” and are “universal” (Sunstein 2005, p. 35).

Whether people are systematically ignorant of the benefits that are forgone as a result of potentially hazardous courses of action can be questioned—as further insights from the cognitive sciences suggest. Although we may be prone to overreaction in some discrete settings, it remains one of the most well-established findings in modern psychology that we tend to justify and excuse the systems within which we are embedded to the point that we ignore or discount evidence that those systems may be unjust, unsustainable, or otherwise undesirable. Take the case of public reactions in the aftermath of catastrophic events such as Hurricane Katrina, in which stark differences in disaster vulnerability were revealed along racial and economic lines in the city of New Orleans and surrounding areas. The responses of both victims and the general public involved a direct defense of the status quo in terms of victim blaming, stereotyping, and internalization of inequality. Rather than criticizing those in charge—given that five disastrous days passed between the storm and a firm commencement of relief efforts—much blame was addressed toward the victims of Hurricane Katrina (Napier et al. 2006). An impartial look at the situation exposed clear governmental shortcomings such as the failure of local and national government to communicate and to provide for adequate evacuation opportunities. But since this criticism would have called into question the legitimacy of the “system,” and would have highlighted stark racial and economic inequalities within it, both the accounts of journalists and evacuees alike reported—without support (Thevenot and Russell 2005)—that many of the storm’s troubling consequences for the poor and displaced were due to violence and other types of unlawful behavior in the days following the hurricane. Likewise, many commentators and observers appeared to hold flood victims responsible for “choosing” not to evacuate, without assuring themselves that the affected individuals actually received a meaningful opportunity to flee the coming storm.

In such situations, people are motivated to accept and rationalize features of existing social arrangements because they seek consistency in their beliefs (psychologists refer to the process of producing internal consistency of beliefs as dissonance reduction; see Festinger 1957), one of which typically is a belief that the social systems in which they are embedded are basically just. This belief inclines them toward processing new information in a manner that is consonant with their prevailing views about the contexts within which they live (Greenwald 1980). Social psychologists speak of a human proclivity to hold more favorable attitudes toward the “system” than is warranted, rejecting evidence that is inconsistent with these attitudes (Lord, Ross, and Lepper 1979). As in the case of processing scientific knowledge, such “biased assimilation” results in a tendency of people’s beliefs about the world to become more extreme rather than being moderated, if they are confronted with mixed evidence on a subject about which they hold strong feelings (Lord, Ross, and Lepper 1979).

From this finding follows a widespread propensity towards system justification (Jost et al. 2008), in that people “support, defend, and bolster the status quo simply because it exists” (Napier et al. 2006, p. 60). Thus, contrary to what proponents of the behavioral

economic approach contend, the demand for environmental legislation might actually be far too *low*—particularly among those deprived of access to environmental quality—because individuals have adapted to an environmentally unsustainable status quo. Whether a regulatory proposal will be perceived as preserving or as threatening the status quo is often a contingent, context-specific question. In the case of climate change, for instance, the US public has overwhelmingly perceived greenhouse gas *regulations* as posing a threat to life and business as usual, given that the usual policy proposals to address climate change entail raising the cost or otherwise significantly altering energy, transportation, agricultural, water management, and other critical infrastructure systems (Feygina, Jost, and Goldsmith 2010). It remains to be seen whether high-profile climate-related disasters such as Hurricane Sandy will alter this dynamic, enabling climate change itself to be seen as a threat to the status quo, rather than the regulations that combat climate change.

Criticism of the same kind applies to what has been inferred from the phenomenon of loss aversion. Rather than blinding individuals to the potential benefits of hazardous courses of action, loss aversion may render them relatively unwilling to sacrifice existing benefits voluntarily. The status quo in these cases acts as a reference point that people seek to maintain. Thus, people are willing to tolerate risks that they already tolerate, even though they would not otherwise be willing to incur the same magnitude of risk from a new source (Rachlinski 2000). In the environmental, health, and safety context, most of the time we are compelled to choose whether to incur a loss from the present status quo in order to obtain a future gain. For this reason, loss aversion may put an additional roadblock in the way of society's efforts to avoid disastrous environmental consequences (Rachlinski 2000; chapter 11 by Zamir in this volume).

Similarly, whenever the legal system allocates entitlements, problems like those generated by loss aversion and the status quo bias are likely to be exacerbated. Consider the fact that it has been difficult, if not impossible, in the United States to bring about environmental regulation through tax increases on polluting products or activities. Although even representatives for the automobile industry have recommended gasoline taxes, serious proposals in that vein are virtually nonexistent in US national environmental policy (Pirog 2011). Such political dynamics may well be understood in terms of the endowment effect. People who have been granted an entitlement to pollute the air will inevitably put a lower value on clean air than they would if they had not been given such an entitlement (Kahneman 1992, pp. 304–5 referring to the asymmetric treatment of losses and gains in conjunction with the notion of rights and entitlements as “enhanced loss aversion”). The current price of gasoline marks the status quo—the entitlement—from which deviations are measured. Given these conditions, it should scarcely come as a surprise that efforts to raise the price of gasoline are persistently met with opposition.<sup>1</sup>

<sup>1</sup> Thomas Merrill and David Schizer have offered an ingenious gasoline tax proposal that takes advantage of status quo bias by establishing a price floor keyed to present US gas prices. If market prices drop, the tax kicks in and consumer opposition is presumably lessened by the perception that prices have not risen above what they perceive to be the status quo (Merrill and Schizer 2008).

The status quo bias, loss aversion, and the endowment effect therefore strongly govern the demand for environmental regulation. Contrary to the view held by many advocates of the behavioral approach, this reading of the research suggests that private and public reactions to risks may be predictably biased *against* more stringent regulation compared to the status quo—a reading that again suggests the proregulatory tilt of the precautionary approach may have much to commend it.

## 4 BEHAVIORALISM ALL THE WAY DOWN

---

The preceding section has shown that the diverse findings of the behavioral research are hardly capable of generating an overall prediction about the manner in which people perceive environmental hazards. The mere recognition of cognitive phenomena such as loss aversion, availability, polarization, and so forth does not produce easy or incontestable answers as to whether people take excessive precautions or, on other plausible assumptions, whether they do exactly the opposite (Hanson and Kysar 1999; Korobkin 2009). Yet the focus by behavioral economics researchers on the *micro level* of human decision-making has put scholars in a position to continue to deploy the conventional model of economic analysis with only the fixed additional input of one or more cognitive biases (Hanson and Kysar 2001, p. 259). This approach falls short, however, of recognizing the manner in which human behavior exists as a function of social, cultural, and political forces at the *macro level*. It is telling that the behavioral economic literature to date has mostly focused on the findings of cognitive psychology, with its emphasis on the individual and on perceived departures from the rational utility-maximizing ideal, rather than on social psychology, with its attempt to understand humans in their full socially embedded complexity. As economists and social psychologists have shown, individuals adapt to the role expectations that are projected onto them (Akerlof and Kranton 2010), to their social and cultural environment (Dahrendorf 1973), to their economic status (Veblen 1994), to political processes (Elster 1983), and so forth. Analysts must therefore study cognitive biases and other behavioral findings not merely as exogenous factors to be added to the basic rational actor model, but also as endogenous forces that may complicate fundamentally the goal of using law and policy to enhance the satisfaction of individual preferences (Hanson and Kysar 1999; Hanson and Kysar 2001; Kuenzler 2008; Sunstein 1993). To put the point succinctly, the correctives prescribed by cognitive psychology may only become the independent variables studied through social psychology.

If one takes this kind of endogeneity seriously, then what results is an entirely different conception of policy analysis than the one that underpins conventional welfare economics, which relies on an aggregation of preexisting preferences to guide societal trade-off decision-making. Preferences that are, at least in part, a function of social and political processes can no longer serve as an independent yardstick for policy decisions. Our notions of preferences are not fully separable from context because our

social, political, and cultural environment works in part to construct our very notions of what we prefer. As Jon Hanson and David Yosifon have cogently stated, “The temptation to see preferences first and behavior as the consequence is widely shared. When we act, we need to believe that there exists an attitude or preference behind that act—our motive for coherence creates a need for reasons. But very often reasons do not precede behavior. So, we are motivated to concoct them—so motivated that we will often alter pre-existing attitudes” (Hanson and Yosifon 2004, p. 122).

The phenomenon of cognitive dissonance, understood by psychologists to be a fundamental feature of human cognition, lies behind this reversal of reason and behavior (Steele 1988). In the environmental regulatory context, we face a contradiction between our beliefs—that everything will somehow work out, that the Antarctic ice shelves will not break off into the ocean, and that our coastlines will not be inundated—and all the evidence demonstrating that these calamities are already underway and that we are on a path toward irreversible climate change. This contradiction creates within us the tempestuous and disturbing state of cognitive dissonance. In order to alleviate our psychological burden we respond intuitively in two ways: first, by reducing the sources of conflict, and second, by ignoring, discarding, or deprecating new information that would increase dissonance. As the classic psychological researcher Leon Festinger observed, these propensities in individuals may attain mass acceptance, boosting a catchism of flawed beliefs. If everyone else thinks the same way, it is much easier to filter out conflicting information. Hence, in order to uphold our personal belief that disastrous climate change will somehow be avoided, we downplay or neglect information that highlights how dire the situation is and we express naive optimism about progress toward another solution (Festinger 1954).

Environmental regulation, seen from this light, can neither be a straightforward effort to maximize existing preferences, nor may it fall prey to the idea that preferences and behaviors are shaped (exclusively) through attitudes and beliefs (the latter idea tends to prevail, however, in the social norms literature; see for an excellent elaboration on this Lane 2012). Rather, it must reflect a choice between competing preference orderings, which coexist within many individuals and which, at least in part, can be enabled or disabled by altering the relevant decision-making context (as neither set of behaviors can actually be said to reveal “true” preferences) (Kysar 2004). As a result, environmental policymakers must realize that preferences (as well as attitudes and beliefs) lose their traditional status as the dominant yardstick for measuring welfare. They can no longer play an exclusive role as goals, but rather acquire an instrumental meaning: They are to be viewed as tools that can be more or less effective for attaining a particular outcome that is justified by *other* normative criteria. As such, they must remain open to critical reflection, deliberation, and management. In turn, laws must be constructed such that they reflect moral and political discourses that directly engage people in reason-giving, norm-creation, and collective self-determination (Lewinsohn-Zamir 2003).

Ultimately, of course, this conclusion will no longer allow us to judge policies on traditional preferentialist efficiency grounds, which must be seen, instead, as a program

for favoring one set of preferences over another. At times, behavioral law and economic scholars seem to be well aware that their prescriptions depend on an unstable normative framework in this manner. For instance, Thaler and Sunstein defend their brand of libertarian paternalism on the ground that “it is legitimate for choice architects to try to influence people’s behavior in order to make their lives longer, healthier, and better” (Thaler and Sunstein 2009, p. 5). This defense seems compelling only because “longer, healthier and better” lives are essentially uncontested objectives that do not open out into a fully-fledged policy conception, especially so long as “better” remains undefined. In more complicated cases it becomes less clear what values a given regulatory policy should be based upon, what risks a particular legal order might tolerate, or why certain default rules should be better than others (Schlag 2010; Eidenmüller 2011). This becomes even more explicit when we look at the traditional economic measure of value, willingness to pay (WTP). WTP rests on a series of normative assumptions with respect to the group of people who are entitled to take part in decisions regarding nature. In the environmental regulatory context, WTP typically ignores rights and preferences of future generations because future generations are not entitled to bid on present-day resources (on the related issue of discounting within CBA, see Kysar 2007). More broadly, WTP will also produce greater expenditures to protect the wealthy than the poor because people with more income and wealth are able to pay more for equivalent goods than people who own and make comparatively less (Schmid 1989). But the problems associated with WTP evaluations of regulatory costs and benefits run even deeper. As the behavioral research has shown, an individual’s WTP to reduce or to avoid environmental, health, or safety risks is often set collectively rather than individually, and unavoidably so. That is, WTP typically rests on the holdings and actions of peers—and not just on the decline or increase in the amounts of goods and services an individual is able to buy as a result of more or less regulation (Easterlin 1974; Frank 1997; Frank and Sunstein 2001). Specifically, an individual’s decline in living standards *relative to others* will diminish his or her WTP for improvements in, say, better quality of air or water. But if *everyone* has to give up the same amount of other goods they can buy as a result of regulation, people will be willing to pay more for the benefits that environmental regulation intends to produce, because they will be able to maintain their relative status in the positional hierarchy.

The frame of reference within which our own valuation network provides us with information about how to determine the overall value of specific risks and rewards should thus be no less a matter of public concern than is, for instance, the quality of our lives or the health of our environment. What is more, if a person’s own evaluation is significantly affected by the evaluation of others—if our self-conception is (at least partially) a function of contextual forces of this general sort—then the frames of libertarian paternalism or libertarian welfarism seem unduly narrow. Perhaps one of the most compelling reasons why we nonetheless continue to engage in the sort of partial equilibrium analysis that libertarian paternalists and welfarists support comes from a cognitive biasing tendency that social psychologists have fittingly termed the fundamental attribution error. As Lee Ross and Richard E. Nisbett have written, “when

people are called upon to interpret the events that unfold around them, they tend to overlook or to make insufficient allowance for situational influences. [They thus fail to] recognize the importance of subjective interpretation, that is, to realize the extent to which behavior can be predicted and understood only in light of the actor's own construal of the situation providing the context for such behavior. . . . They account for past actions and outcomes, and make predictions about future actions and outcomes, in terms of the person—or more specifically, in terms of the presumed personality traits or other distinctive enduring personal dispositions” (Ross and Nisbett 1991, p. 90). By treating background structural conditions in an economy and society, including the whole stock of critical natural capital and ecosystem services across space and time as less important, and the function of salient dispositions as more important, libertarian paternalists and welfarists can indeed account for the way in which cognitive biases affect specific actions and valuations. But doing so leads them to overlook the manner in which the background itself shapes people's subjective construals and helps to constitute their dispositions.

In the environmental regulatory context, we might thus conclude that the CBA framework is not rich enough to deliver a satisfying formula to protect those resources of nature that are indispensable for preserving human welfare, and for which replacement, beyond a certain threshold, may be difficult or impossible to achieve. As Joshua Farley has explained, “when we increase the harvest of living ecosystem structure, we often diminish its capacity to reproduce itself, reducing not only the potential supply in the future, but also the ecosystem services that structure would otherwise provide” (Farley 2008, p. 1402). Any price-based valuation of our environment therefore runs the risk of exhausting environmentally significant goods, irrespective of whether there are substitutes or not. In Daniel Bromley's rendering of this account, he characterizes such exploitation in terms of intertemporal externalities because future generations cannot bid on present-day resources. Rather, the interests of future generations can only be preserved by an entitlement structure that imposes on present-day generations a duty to consider the interests of the future. In the environmental, health, and safety context, future generations should therefore be endowed with a legal right to the preservation of certain natural goods and services in order to ensure ecological sustainability (Bromley 1995). Such an alternative entitlement structure would then allow marginalist economic analysis to proceed against a general equilibrium backdrop that better ensures environmental goals (Kysar 2011). Moreover, the backdrop would help to ensure that individuals view intergenerational equity and environmental sustainability as collective goals worth valuing.

The need for such an alternative structure is underscored by the fact that many contributions of ecosystems to our well-being are essentially beneath our perception (Vatn and Bromley 1994; McKibbin and Wilcoxon 2002). The inherent complexity, dynamism, and nonlinear behavior that characterize many components of ecological-economic systems may lead to reasonably linear changes in the value of ecosystem services over a particular range but highly nonlinear transformations over another. The course of history is quite frequently punctuated by such seemingly inexplicable upheavals.



Compounding our uncertainty, we may expect long periods of relative and gradual change to be disrupted at intervals by overwhelming events that totally reshape an ecological landscape (Farber, Costanza, and Wilson 2002; Limburg et al. 2002). By the time we are aware of the disruption, its consequences may be irreversible. Like a keg with its tap at the bottom, marginal methods of valuation might reveal a steady flow while failing to indicate that the overall stock within is rapidly dwindling and that our party may soon come to an end. The behavioral approach to regulation thus fails to appreciate that only after certain sustainability requirements have *first* been met can we sensibly use marginal analysis to improve efficiency. Our tendency to perceive only the foreground blinds us to reason about the background—the environmental supply curve rather than the apparent human demand curve (Farley 2008).

If under these conditions normative policy analysis has to move beyond preferences to focus on the factors governing their formation, this calls for a shift of attention towards the social scripts, actors, institutions, and other components of our world that shape preferences, attitudes, and beliefs. A policy that takes these components into account is not content to simply give people “what they want”—because there is no clear way for the law to respect what people really want (Hill 2008; Mitchell 2005)—but rather focuses on the process of how individuals identify what they want (Tribe 1972; Frankfurt 1971). If as a result a consensus emerges, such a consensus cannot be defined on the basis of whatever the individuals happen to prefer at a given point in time. Instead, such a consensus must be understood in light of what *makes* people prefer the things they prefer. It then becomes clear that environmental policy must be based on a completely different (and more terrifying) understanding of “freedom” than that on which the libertarian paternalism or welfarism paradigms are based. Laurence Tribe once expressed this view very succinctly as follows: “[T]o be free is to choose what we shall value” (Tribe 1985, p. 617).

On the basis of such a discursive policy concept, aspects come to the fore that are plainly neglected by the behavioral law and economic literature at present. In addition to reinvigorated debate over whether alternative concepts provide a superior substantive definition of welfare to preferentialism (Zamir and Medina 2010), what also becomes essential are the structural and procedural rules for legitimizing deliberative processes (Habermas 1987). When understood and respected as political subjects, individuals are seen as capable of adapting their preferences and their understandings of moral and ethical responsibilities, rather than simply having those positions be viewed either as sovereign inputs to policymaking or as irrational detritus (Kysar 2004; Kuenzler 2012b). Issues of that sort depend not on an aggregation of preexisting individual preferences or on technocratic “laundering” of preferences thought to be ill-founded, but rather on the collective commitment of the citizenry to apprehend some shared ideal and translate it into public policy. As Amartya Sen has written with respect to policy goals more generally, the environmental threats that we face depend on “value formation, related to public discussions, both for their influence on individual behavior and for bringing about policy changes through the political process” (Sen 1995, p. 17; see also Knight 1947, p. 280, noting that values are “established, validated



[and] recognized through discussion, an activity which is at once social, intellectual, and creative”). In analyzing them “we need to depart both from the assumption of given preferences. . . and from the presumption that people are narrowly self-interested homo oeconomicus” (Sen 1995, pp. 17–18).

What might such an expanded conception of public policy analysis require us to consider? As we have seen, from a microeconomic perspective, all values may be reduced to a single metric even though in real life people sometimes reject certain trade-offs because they believe that giving the relevant comparisons serious consideration would compromise their self-conception and their social identities as ethical beings (Fiske and Tetlock 1997). Such expressions of incommensurability suggests that we have to come to understand the “utility space” that is governed by such norms as communal sharing, authority ranking, and equality matching as distinct from the “utility space” governed by monetary valuation and market exchange. By gaining presence and influence in the former sphere, the tools of monetary valuation, market pricing, and the like may create an increasingly hospitable environment for their own reproduction in the minds of human actors.

Social psychologists have come to much the same conclusions about our capacity to stimulate an individual’s ability to feel the need to act. As the research on numerical representations of saved lives above has shown, not only do the numbers that represent statistics often fail to motivate individual action but the mere presence of statistics can significantly diminish the attention necessary to establish an essential stimulus to act. A case in point is Deborah Small, George Loewenstein, and Paul Slovic’s study on the impact of deliberative thought on donations to *identifiable* versus *statistical* victims. In their experiment, subjects who were given a chance to contribute some of their earnings to an identified, seven-year-old African girl named Rokia gave away more than twice the amount given by another group of subjects that was asked to give to the same organization to save millions of Africans from hunger. Furthermore, adding a display of statistical information to those subjects who were asked to give to Rokia substantially *lessened* the amount of their donations (Small, Loewenstein, and Slovic 2007). As commentators conclude from these findings, “To the extent that valuation of life-saving depends on feelings driven by attention or imagery, it. . . is greatest at  $N=1$  but begins to decline at  $N=2$  and collapses at some higher value of  $N$  that becomes simply ‘a statistic.’ In other words. . . perhaps the ‘blurring’ of individuals begins at two!” (Slovic et al. 2013, pp. 132–33). Subsequent studies support the hypothesis that numerical representations of lives may “numb,” or even “turn off” our feelings of sympathy and our desires to act. Contributions to identifiable (but not statistical) victims tend to be considerably lower when individuals are instructed to do simple arithmetic calculations before they are given an opportunity to donate; donations, on the other hand, are typically higher if participants are first primed to feel (by, for instance, describing their emotions when they hear the word “baby”) (Small, Loewenstein, and Slovic 2007).

These and other observations indicate that people’s intuitive moral responses to particular cases may originate from sources far detached from (one set of) their values and

may sometimes fail to reflect these values, and often even point in opposite directions (Unger 1996, p. 11). They may lead people to behave in ways that are in conflict with the principles they would hold after careful consideration. Thus, a deliberative public policy must be designed broadly, by establishing not only the prerequisites for a shared dialogue among different individuals but also, significantly, by generating conditions that provoke critical thinking and result in deeply reasoned judgments even before an open dialogue begins. Such deliberative processes can be structured so as to counteract the rationality deficits of those involved in the proceedings in a targeted way (Elster 1983; van Aaken 2007; Eidenmüller 2011). Such a concept is even conceivable in the realm of market regulation (Kuenzler 2012a).

Consider the alternative conditions established by a public policy apparatus too heavily fixated on welfare economics and CBA. In this society, the economic approach begins increasingly to shape and create the culture within which its value is produced and, for that very reason, to perpetuate the conditions of its own flourishing. After all, as a number of influential studies have shown, where public spirit prevails, using *price* incentives to muster support for the construction of a socially desirable enterprise may induce people to be less helpful to others and to prefer to act selfishly. To mention but one example, Bruno Frey and Felix Oberholzer-Gee discovered evidence of such motivational competition when they evaluated the opinions of Swiss citizens with respect to the siting of nuclear waste facilities. When asked whether they would be willing to have a waste dump in their community, 51% of respondents agreed although they were aware of the potential dangers and impacts on their property values. When other people were asked to have the waste dumps in their communities in return for an annual payment equivalent to six weeks' worth of an average Swiss salary, only 25% of respondents said yes (Frey and Oberholzer-Gee 1997). Adding the financial incentive thus resulted in only half as much acceptance although the respondents now had two reasons to say yes (obligations as citizens *and* financial incentives). As Barry Schwartz has explained, in many cases "the introduction of fines or the offer of compensation relocates the activities from a different social sphere into the market sphere" (Schwartz 2012, p. 171). Indeed, it often takes little effort to shift the character of an activity in that direction. In other studies, the mere mention of monetary payment was sufficient to switch the perceived relationship from a social one to a market one (Heyman and Ariely 2004).

Thus, market relationships are a sort of "ecology of reproduction" and hence are in competition with other forms of human relationships. This ecology has important disciplinary effects. By creating and recreating expectations about what one should value and not value, about what one should reflect and not reflect, market relationships normalize certain practices of thought and action and make others deviant. Turning environmental issues into market pricing matters alters our evaluative frame, irrespective of whether we are asked how much we would be willing to pay to avoid pollution or how much we would be willing to accept to endure it. Thus, the systematic use of CBA to evaluate the environment—rather than thinking explicitly about knotty issues of value incommensurability or intergenerational sustainability—already marks a regulatory approach as a certain type of institutional setting that promotes a particular

view of the environment and reflects an insufficient desire to leave space for other types of assessments. Hence, it matters whether different activities are located on one or the other terrain of our interactions with each other, precisely because our very notions of the environment may become infected with and even constituted by those relationships. Once the move to locate an issue on a common monetary scale is made, the issue in question is placed in the market pricing system of social relations (Schwartz 2012). This may reduce the likelihood of cooperation—simply because an issue for which people would otherwise be willing to sacrifice is turned into a market pricing matter. As a result, individuals think differently about the issue and make different decisions. Instead of merely measuring people’s preferences, CBA may actually mold their preferences by predetermining the *standard* of evaluation, which in turn helps to predetermine the *outcome* of the evaluation (Schwartz 2012).

## 5 CONCLUSION

---

As Russell Korobkin has written, “In the wake of the behavioral revolution, law and economics cannot progress as a normative discipline without a richer theory of welfare than it now possesses” (Korobkin 2011, p. 1664). The behavioralist approach to environmental, health, and safety regulation makes this point plain. Through a gradual development, the structure of risk regulation in the United States and elsewhere has come to embody a centralized and proceduralized mode of risk assessment and CBA, with an emphasis on the use of market-based mechanisms in the event that regulation is justified by these analytical tools. Behavioral legal economists have sought to support this structure and enrich it through an emphasis on selective insights from cognitive psychology and other social science literatures. As this chapter has argued, however, the mainstream behavioral law and economics paradigm fails to adequately account for the endogenous effects of cognitive biases, from the perception and cultural encoding of risk by experts and laypeople to the selection and choice of regulatory instruments by policymakers.

Endogeneity—the most theoretically disruptive and important implication of behavioralism—implies that individual expressions of value and risk may be influenced by social conceptions of valuation and risk assessment, even when the latter are designed merely to measure the former without also affecting them. For instance, if we embrace the view that monetary valuations of individual preferences are all that matters in the determination of public policy, then we may be led to create political structures that cater to self-interest and to those human values that are readily reducible to monetary terms. As a result, we may shape a society in which the assumption that self-interest and monetary valuation are dominant becomes a self-fulfilling truth. It is not hard to imagine that if environmental regulation becomes dominated by market pricing, people will come to understand our environmental challenges as a matter of optimizing market pricing relations. Far from being an exercise in speaking truth

passively and objectively to regulatory decision-makers, then, CBA may instead be a longer-term project that actively creates new truths in a society.

From the constructedness of these categories it follows that some of the concerns people have as citizens cannot in principle be expressed in their roles as consumers, but must be expressed through their political relations with other citizens. Broadly speaking, consumers act individually, taking the background social relations of their interactions for granted and generally assuming an instrumental attitude toward those relations. Citizens act collectively, taking their social relations as an intrinsic object of concern and self-conscious construction. Because these relations are constituted by shared legal, ethical, and social norms, people can reform them only through collective action (Anderson 1993; Lewinsohn-Zamir 1998). Through the use of democratic institutions we enable people to express certain *kinds* of valuations that can be expressed only in nonmarket social relations. Which institution—the state or the market—we should use to govern outcomes concerning environmental dilemmas is therefore not strictly a question of efficiency. Rather, the choice of institutions should be determined by the kind of goods at stake, the kinds of concerns people have with respect to them, and—critically—the kinds of values we wish people to embody.

If we want to prevent the redefinition of our environment within the instrumentalist language of economy and administration, we should therefore cease conceiving of people's self-interest in particular settings as a *general* account of human nature. We should acknowledge instead that even though the idea of self-interested behavior captures a significant social phenomenon, this phenomenon itself is an invention; a creation that, once embraced, reshapes one social institution after another such that dramatic changes in behavior follow (Schwartz 2012). As a consequence of the dynamics that lie behind purportedly impartial accounts of human behavior, the discipline of economics itself does not inevitably have to be a theory of prediction and control where government can achieve its ends only by providing a proper set of incentives for the desired behavior. A better answer to the environmental problems that we need to solve might instead be found in a second, more generative role for economic theory. By this we mean its role in creating values and behaviors as opposed to merely predicting or controlling them (Kuenzler 2012a).

## REFERENCES

- Akerlof, G. A., and R. E. Kranton. 2010. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton, NJ: Princeton University Press.
- Adler, M. D., and E. A. Posner. 2006. *New Foundations of Cost-Benefit Analysis*. Cambridge, MA: Harvard University Press.
- Anderson, E. 1993. *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.
- Ayres, I., S. Raseman, and A. Shih. 2009. Evidence from Two Large Field Experiments That Peer Comparison Feedback Can Reduce Residential Energy Usage. NBER Working Paper No. 15386. Accessed September 26, 2012. <http://www.nber.org/papers/w15386>.

- Baron, R. S., S. I. Hoppe, C. F. Kao, B. Brunsman, B. Linneweh, and D. Rogers. 1996. Social Corroboration and Opinion Extremity. *Journal of Experimental Social Psychology* 32:537–60.
- Becker, G. 1976. *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press.
- Berman, E., and L. T. M. Bui Linda. 2001. Environmental Regulation and Productivity: Evidence from Oil Refineries. *Review of Economics and Statistics* 83:498–510.
- Bezdeka, R. H., R. M. Wendlinga, and P. DiPerna. 2008. Environmental Protection, the Economy, and Jobs: National and Regional Analyses. *Journal of Environmental Management* 86:63–79.
- Boyd, W., D. A. Kysar, and J. J. Rachlinski. 2012. Law, Environment, and the “Non-dismal” Social Sciences. *Annual Review of Law and Social Science* 8:183–211.
- Bromley, D. W. 1995. Property Rights and Natural Resource Damage Assessments. *Ecological Economics* 14:129–35.
- Burrows, P. 1993. Patronising Paternalism. *Oxford Economic Papers* 45:542–72.
- Camerer, C., S. Issacharoff, G. Loewenstein, T. O’Donoghue, and M. Rabin. 2003. Regulation for Conservatives: Behavioral Economics and the Case for Asymmetric Paternalism. *University of Pennsylvania Law Review* 151:1211–54.
- Costa, D. L., and M. E. Kahn. 2010. Energy Conservation “Nudges” and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment. NBER Working Paper No. 15939. Accessed January 22, 2013. <http://www.nber.org/papers/w15939>.
- Dahrendorf, R. 1973. *Homo Sociologicus*. London: Routledge.
- Dana, D. A. 2003. A Behavioral Economic Defense of the Precautionary Principle. *Northwestern University Law Review* 97:1315–46.
- Easterlin, R. A. 1974. Does Economic Growth Improve the Human Lot? Some Empirical Evidence. In *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, P. A. David and M. W. Reder. New York: Academic, 89–125.
- Eidenmüller, H. 2011. Liberaler Paternalismus. *Juristen Zeitung* 17:814–21.
- Elster, J. 1983. *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge University Press.
- Executive Order 12291. February 19, 1981. 46 Fed. Reg. 13193.
- Farber, S. C., R. Costanza, and M. A. Wilson. 2002. Economic and Ecological Concepts for Valuing Ecosystem Services. *Ecological Economics* 41:375–92.
- Farley, J. 2008. The Role of Prices in Conserving Critical Natural Capital. *Conservation Biology* 22:1399–408.
- Festinger, L. 1954. A Theory of Social Comparison Processes. *Human Relations* 7:117–40.
- Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
- Fetherstonhaugh, D., P. Slovic, S. M. Johnson, and J. Friedrich. 1997. Insensitivity to the Value of Human Life: A Study of Psychophysical Numbing. *Journal of Risk and Uncertainty* 14:283–300.
- Feygina, I., J. T. Jost, and R. E. Goldsmith. 2010. System Justification, the Denial of Global Warming, and the Possibility of “System-Sanctioned Change.” *Personality and Social Psychology Bulletin* 36:326–38.
- Finucane, M. L., A. Alhakami, P. Slovic, and S. M. Johnson. 2000. The Affect Heuristic in Judgments of Risks and Benefits. *Journal of Behavioral Decision Making* 13:1–17.

- Fischhoff, B., P. Slovic, S. Lichtenstein, S. Read, and B. Combs. 1978. How Safe Is Safe Enough? A Psychometric Study of Attitudes towards Technological Risks and Benefits. In *The Perception of Risk*, P. Slovic. London: Earthscan, 137–53.
- Fiske, A. P., and P. E. Tetlock. 1997. Taboo Trade-Offs: Reactions to Transactions That Transgress the Spheres of Justice. *Political Psychology* 18:255–97.
- Frank, R. H. 1997. The Frame of Reference as a Public Good. *Economic Journal* 107:1832–47.
- Frank, R. H., and C. R. Sunstein. 2001. Cost-Benefit Analysis and Relative Position. *University of Chicago Law Review* 68:323–47.
- Frankfurt, H. G. 1971. Freedom of the Will and the Concept of a Person. *Journal of Philosophy* 68:5–20.
- Frey, B. S., and F. Oberholzer-Gee. 1997. The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out. *American Economic Review* 87:746–55.
- Gayer, T., and W. K. Viscusi. 2012. Overriding Consumer Preferences with Energy Regulations. Accessed October 1, 2012. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2111450](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2111450).
- Gigerenzer, G., P. M. Todd, and the ABC Research Group. 1999. *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.
- Graham, J. D., and J. B. Wiener. 1995. Confronting Risk Tradeoffs. In *Risk v. Risk: Tradeoffs in Protection Health and the Environment*, J. D. Graham and J. B. Wiener. Cambridge, MA: Harvard University Press, 1–41.
- Greenwald, A. G. 1980. The Totalitarian Ego: Fabrication and Revision of Personal History. *American Psychologist* 35:603–18.
- Habermas, J. 1987. *The Theory of Communicative Action*. Boston: Beacon Press.
- Hanson, J. D., and D. A. Kysar. 1999. Taking Behavioralism Seriously. The Problem of Market Manipulation. *New York University Law Review* 74:630–749.
- Hanson, J. D., and D. A. Kysar. 2001. The Joint Failure of Economic Theory and Legal Regulation. In *Smoking, Risk, Perception, and Policy*, P. Slovic. London: Sage, 229–76.
- Hanson, J., and D. Yosifon. 2004. The Situational Character: A Critical Realist Perspective on the Human Animal. *Georgetown Law Journal* 93:1–179.
- Hatfield, E., J. T. Cacioppo, and R. L. Rapson. 1994. *Emotional Contagion*. New York: Cambridge University Press.
- Heath, C. 1996. Do People Prefer to Pass Along Good or Bad News? Valence and Relevance as Predictors of Transmission Propensity. *Organizational Behavior and Human Decision Process* 68:79–94.
- Heinzerling, L. 1998. Regulatory Costs of Mythic Proportions. *Yale Law Journal* 107:1981–2070.
- Heyman, J., and D. Ariely. 2004. Effort for Payment: A Tale of Two Markets. *Psychological Science* 15:787–93.
- Hill, C. A. 2008. The Rationality of Preference Construction (and the Irrationality of Rational Choice). *Minnesota Journal of Law, Science, and Technology* 9:689–742.
- Hill, S. 1992. *Democratic Values and Technological Choices*. Stanford, CA: Stanford University Press.
- Jacob, K., S. Weiland, J. Ferretti, D. Wascher, and D. Chodorowska. 2011. Integrating the Environment in Regulatory Impact Assessments. OEDC. Accessed May 4, 2013. <http://www.oecd.org/gov/regulatory-policy/ria.htm>.
- Jolls, C., and C. R. Sunstein. 2006. The Law of Implicit Bias. *California Law Review* 94:969–96.



- Jolls, C., C. R. Sunstein, and R. Thaler. 1998. A Behavioral Approach to Law and Economics. *Stanford Law Review* 50:1471–550.
- Jost, J. T., J. Pietrzak, I. Liviatan, A. N. Mandisodza, and J. L. Napier. 2008. System Justification as Conscious and Nonconscious Goal Pursuit. In *Handbook of Motivation Science*, J. J. Shah and W. Gardner. New York: Guilford, 591–605.
- Kagan, S. 1998. *Normative Ethics*. Boulder, CO: Westview Press.
- Kahan, D. M., and D. Braman. 2003. More Statistics, Less Persuasion: A Cultural Theory of Gun-Risk Perceptions. *University of Pennsylvania Law Review* 151:1291–328.
- Kahan, D. M., P. Slovic, D. Braman, and J. Gastil. 2006. Fear of Democracy: A Cultural Evaluation of Sunstein on Risk. *Harvard Law Review* 119:1071–109.
- Kahneman, D. 1992. Reference Points, Anchors, Norms, and Mixed Feelings. *Organizational Behavior and Human Decision Processes* 51:296–312.
- Kahneman, D. 2003. Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review* 93:1449–75.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1990. Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy* 98:1325–48.
- Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47:263–92.
- Knight, F. 1947. *Freedom and Reform: Essays in Economic and Social Philosophy*. New York: Harper.
- Kochi, I., B. Hubbell, and R. Kramer. 2006. An Empirical Bayes Approach to Combining and Comparing Estimates of the Value of a Statistical Life for Environmental Policy Analysis. *Environmental and Resource Economics* 34:385–406.
- Korobkin, R. 2009. Libertarian Welfarism. *California Law Review* 97:1651–86.
- Korobkin, R. 2011. What Comes after Victory for Behavioral Law and Economics? *University of Illinois Law Review* 2011:1653–74.
- Krier, J. E. 1994. The End of the World News. *Loyola of Los Angeles Law Review* 27:851–66.
- Kuenzler, A. 2008. *Effizienz oder Wettbewerbsfreiheit? Zur Frage nach den Aufgaben des Rechts gegen private Wettbewerbsbeschränkungen*. Tübingen: Mohr Siebeck.
- Kuenzler, A. 2012a. Economic Content of Competition Law: The Point of Regulating Preferences. In *The Goals of Competition Law*, D. Zimmer. Northampton, MA: Edward Elgar, 182–213.
- Kuenzler, A. 2012b. The Jurisprudence of Welfare Maximization: Embracing Complexity and Uncertainty in the Environmental, Health, and Safety Context. *Archives for Philosophy of Law and Social Philosophy* 98:1–18.
- Kuran, T., and C. R. Sunstein. 1999. Availability Cascades and Risk Regulation. *Stanford Law Review* 51:683–768.
- Kysar, D. A. 2003. The Expectations of Consumers. *Columbia Law Review* 103:1700–90.
- Kysar, D. A. 2004. Climate Change, Cultural Transformation, and Comprehensive Rationality. *Boston College Environmental Affairs Law Review* 31:555–90.
- Kysar, D. A. 2004. Preferences for Processes: The Process/Product Distinction and the Regulation of Consumer Choice. *Harvard Law Review* 118:526–642.
- Kysar, D. A. 2007. Discounting. . . on Stilts. *University of Chicago Law Review* 74:119–38.
- Kysar, D. A. 2009. Fish Tales. In *Reforming Regulatory Impact Analysis*, W. Harrington Winston, L. Heinzerling, and R. D. Morgenstern. Washington, DC: Meadows, 150–214.
- Kysar, D. A. 2011. Politics by Other Meanings: A Comment on “Retaking Rationality Two Years Later.” *Houston Law Review* 48:43–78.



- Kysar, D. A., P. Ayton, R. H. Frank, B. S. Frey, G. Gigerenzer, P. W. Glimcher, R. Korobkin, D. C. Langevoort, and S. Magen. 2005. Group Report: Are Heuristics a Problem or a Solution? In *Heuristics and the Law*, G. Gigerenzer and C. Engel. Cambridge, MA: MIT Press, 103–39.
- Lane, M. 2012. *Eco-republic: What the Ancients Can Teach Us about Ethics, Virtue, and Sustainable Living*. Princeton, NJ: Princeton University Press.
- Lazarus, R. 2004. *The Making of Environmental Law*. Chicago: University of Chicago Press.
- Lewinsohn-Zamir, D. 1998. Consumer Preferences, Citizen Preferences, and the Provision of Public Goods. *Yale Law Journal* 108:377–406.
- Lewinsohn-Zamir, D. 2003. The Objectivity of Well-Being and the Objectives of Property Law. *New York University Law Review* 78:1669–754.
- Limburg, K. E., R. V. O'Neill, R. Costanza, and S. C. Farber. 2002. Complex Systems and Valuation. *Ecological Economics* 41:409–20.
- Livermore, M. A., and R. L. Revesz. 2011. Retaking Rationality Two Years Later. *Houston Law Review* 48:1–41.
- Livermore, M. A., and R. L. Revesz. 2013. *The Globalization of Cost-Benefit Analysis in Environmental Policy*. New York: Oxford University Press.
- Loewenstein, G., and E. Haisley. 2008. The Economist as Therapist: Methodological Ramifications of “Light” Paternalism. In *The Foundations of Positive and Normative Economics: A Handbook*, A. Caplin and A. Schotter. New York: Oxford University Press, 210–46.
- Loewenstein, G., and D. Prelec. 1992. Anomalies in Intertemporal Choice: Evidence and Interpretation. In *Choice over Time*, G. Loewenstein and J. Elster. New York: Russell Sage, 119–45.
- Loewenstein, G., E. U. Weber, C. K. Hsee, and N. Welch. 2001. Risk as Feelings. *Psychological Bulletin* 127:267–86.
- Lord, C. G., L. Ross, and M. R. Lepper. 1979. Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology* 37:2098–109.
- McGarity, T. O. 2004. The Goals of Environmental Legislation. *Boston College Environmental Affairs Review* 31:529–54.
- McGarity, T. O. 2012. Administrative Law as Blood Sport: Policy Erosion in a Highly Partisan Age. *Duke Law Journal* 61:1671–762.
- McHughen, A. 2000. *Pandora's Picnic Basket*. New York: Oxford University Press.
- McKibbin, W. J., and P. J. Wilcoxon. 2002. The Role of Economics in Climate Change Policy. *Journal of Economic Perspectives* 16:107–29.
- Merrill, T., and D. Schizer. 2008. A Standby Levy on Gasoline. Accessed January 25, 2012. [http://www.law.columbia.edu/null/David+Schizer\\_A+Stanby+Levy+on+Gasoline?exclude=filemgr.download&file\\_id=16660&showthumb=0](http://www.law.columbia.edu/null/David+Schizer_A+Stanby+Levy+on+Gasoline?exclude=filemgr.download&file_id=16660&showthumb=0).
- Mitchell, G. 2005. Libertarian Paternalism Is an Oxymoron. *Northwestern University Law Review* 99:1245–77.
- Morgan, M. G. 1993. Risk Analysis and Management. *Scientific American* 32:32–41.
- Morgenstern, R. D., W. A. Pizer, and J.-S. Shih. 2002. Jobs versus the Environment: An Industry-Level Perspective. *Journal of Environmental Economics and Management* 43:412–36.
- Napier, J. L., A. N. Mandisodza, S. M. Andersen, and J. T. Jost. 2006. System Justification in Responding to the Poor and Displaced in the Aftermath of Hurricane Katrina. *Analyses of Social Issues and Public Policy* 6:57–73.

- OECD. 1997. *Regulatory Impact Analysis: Best Practices in OECD Countries*. Accessed May 4, 2013. <http://www.oecd.org/gov/regulatory-policy/35258828.pdf>.
- OIRA. 2009. Office of Information and Regulatory Affairs, Office of Management and Budget, Report to Congress on the Benefits and Costs of Federal Regulations and Unfunded Mandates on State, Local, and Tribal Entities.
- Pirog, R. 2011. The Role of Federal Gasoline Excise Taxes in Public Policy. CRS Report for Congress, August 16. Accessed October 2, 2012. <http://www.nationalaglawcenter.org/assets/crs/R40808.pdf>.
- Rachlinski, J. J. 2000. The Psychology of Global Climate Change. *University of Illinois Law Review* 2000:299–320.
- Regan, D. H. 1983. Paternalism, Freedom, Identity, and Commitment. In *Paternalism*, R. Sartorius. Minneapolis: University of Minnesota Press, 113–38.
- Revesz, R. L. 1999. Environmental Regulation, Cost-Benefit Analysis, and the Discounting of Human Lives. *Columbia Law Review* 99:941–1017.
- Revesz, R. L., and M. A. Livermore. 2008. *Retaking Rationality: How Cost-Benefit Analysis Can Better Protect the Environment and Our Health*. New York: Oxford University Press.
- Ross, L., and R. E. Nisbett. 1991. *The Person and the Situation: Perspectives of Social Psychology*. Philadelphia: Temple University Press.
- Rottenstreich, Y., and C. K. Hsee. 2001. Money, Kisses, and Electric Shocks: On the Affective Psychology of Risk. *Psychological Science* 12:185–89.
- Rozin, P. 2001. Technological Stigma: Some Perspectives from the Study of Contagion. In *Risk, Media, and Stigma: Understanding Public Challenges to Modern Science and Technology*, P. Slovic, J. Flynn, and H. Kunreuther. London: Earthscan, 31–39.
- Schlag, P. 2010. Nudge, Choice Architecture, and Libertarian Paternalism. *Michigan Law Review* 108:913–24.
- Schmid, A. A. 1989. *Benefit-Cost Analysis: A Political Economy Approach*. Boulder, CO: Westview Press.
- Schulze, W., K. Messer, K. Hackett, T. Cameron, G. McClelland. 2004. Stigma: The Psychology and Economics of Superfund. Accessed October 2, 2012. [http://yosemite.epa.gov/ee/epa/eam.nsf/vwAN/EE-0486-01.pdf/\\$file/EE-0486-01.pdf](http://yosemite.epa.gov/ee/epa/eam.nsf/vwAN/EE-0486-01.pdf/$file/EE-0486-01.pdf).
- Schwartz, B. 2012. Crowding Out Morality: How the Ideology of Self-Interest Can Be Self-Fulfilling. In *Ideology, Psychology, and Law*, J. D. Hanson. New York: Oxford University Press, 160–84.
- Sen, A. 1995. Rationality and Social Choice. *American Economic Review* 85:1–24.
- Sen, A. 2000. The Discipline of Cost-Benefit Analysis. *Journal of Legal Studies* 29:931–52.
- Shelley, M. K. 1994. Gain/Loss Asymmetry in Risky Intertemporal Choice. *Organisational Behavior and Human Decision Processes* 59:124–59.
- Slovic, P. 2000. *The Perception of Risk*. London: Routledge.
- Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor. 2002. The Affect Heuristic. In *Heuristics and Biases: The Psychology of Intuitive Judgment*, T. Gilovich, D. Griffin, and D. Kahneman. New York: Cambridge University Press, 397–420.
- Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor. 2004. Risk as Analysis and Risk as Feelings: Some thoughts about Affect, Reason, Risk, and Rationality. *Risk Analysis* 24:311–22.
- Slovic, P., D. Zionts, A. K. Woods, R. Goodman, and D. Jinks. 2013. Psychic Numbing and Mass Atrocity. In *The Behavioral Foundations of Public Policy*, E. Shafir. New York: Oxford University Press, 126–42.

- Small, D. A., G. Loewenstein, and P. Slovic. 2007. Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims. *Organizational Behavior and Human Decision Processes* 102:143–53.
- Steele, C. M. 1988. The Psychology of Self-Affirmation: Sustaining the Integrity of the Self. In *Advances in Experimental Social Psychology* 21, L. Berkowitz. London: Elsevier, 261–302.
- Sunstein, C. R. 1993. Endogenous Preferences, Environmental Law. *Journal of Legal Studies* 22:217–54.
- Sunstein, C. R. 2000. Cognition and Cost-Benefit Analysis. *Journal of Legal Studies* 29:1059–104.
- Sunstein, C. R. 2002. *The Cost-Benefit State: The Future of Regulatory Protection*. Chicago: American Bar Association.
- Sunstein, C. R. 2003. Beyond the Precautionary Principle. *University of Pennsylvania Law Review* 151:1003–58.
- Sunstein, C. R. 2005. *Laws of Fear: Beyond the Precautionary Principle*. New York: Cambridge University Press.
- Sunstein, C. R. 2006. The Availability Heuristic, Intuitive Cost-Benefit Analysis, and Climate Change. *Climate Change* 77:195–210.
- Sunstein, C. R. 2011. Empirically Informed Regulation. *University of Chicago Law Review* 78:1349–429.
- Sunstein, C. R., and R. H. Thaler. 2003. Libertarian Paternalism Is Not an Oxymoron. *University of Chicago Law Review* 70:1159–202.
- Thaler, R. H. 1991. The Psychology of Choice and the Assumptions of Economics. In *Quasi Rational Economics*, R. H. Thaler. New York: Russell Sage, 137–66.
- Thaler, R. H., and C. R. Sunstein. 2009. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. London: Penguin.
- Thevenot, D., and G. Russell. 2005. Rumors of Deaths Greatly Exaggerated. *Times Picayune*, September 26. Accessed October 2, 2012. [http://nola.com/newslogs/tporleans/index.ssf?/mtlogs/nolatporleans/archives/2005\\_09\\_26.html#082732](http://nola.com/newslogs/tporleans/index.ssf?/mtlogs/nolatporleans/archives/2005_09_26.html#082732).
- Tribe, L. H. 1972. Policy Science: Analysis or Ideology? *Philosophy and Public Affairs* 2:66–110.
- Tribe, L. H. 1985. Constitutional Calculus: Equal Justice or Economic Efficiency? *Harvard Law Review* 98:592–621.
- Tversky, A., and D. Kahneman. 1982. Judgment under Uncertainty. In *Judgment under Uncertainty: Heuristics and Biases*, D. Kahneman Daniel, P. Slovic, and A. Tversky. New York: Cambridge University Press, 3–21.
- Unger, P. 1996. *Living High and Letting Die: Our Illusion of Innocence*. New York: Oxford University Press.
- Van Aaken, A. 2007. Begrenzte Rationalität und Paternalismusgefahr: Das Prinzip des schonendsten Paternalismus. In *Recht und Verhalten*, C. Engel, M. Englerth, J. Lüdemann, and I. Spieker. Tübingen: Mohr Siebeck, 189–230.
- Vandenbergh, M. P. 2005. Order without Social Norms: How Personal Norm Activation Can Protect the Environment. *Northwestern University Law Review* 99:1101–66.
- Vatn, A., and D. W. Bromley. 1994. Choices without Prices without Apologies. *Journal of Environmental Economics and Management* 26:129–48.
- Veblen, T. 1994. *The Theory of the Leisure Class: An Economic Study of Institutions*. New York: Dover Publications.
- Viscusi, W. K. 1994. Risk-Risk Analysis. *Journal of Risk and Uncertainty* 8:5–17.
- Viscusi, W. K. 2000. The Value of Life in Legal Contexts: Survey and Critique. *American Law and Economics Review* 2:195–222.

- Viscusi, W. K., and J. E. Aldy. 2003. The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World. *Journal of Risk and Uncertainty* 27:5–76.
- Vlek, C., and G. Keren. 1992. Behavioral Decision Theory and Environmental Risk Management: Assessment and Resolution of Four “Survival” Dilemmas. *Acta Psychologica* 80:249–78.
- Zamir, E., and B. Medina. 2010. *Law, Economics, and Morality*. New York: Oxford University Press.

## INDEX OF NAMES

---

- Abadie, A., 280  
Abaluck, J.T., 739  
Abel, R.L., 623  
Abeler, J., 270  
Abraham, K., 491  
Abrams, D., 256  
Ackerman, B.A., 380, 397n13  
Adamowicz, W.L., 305–306, 393  
Adams, R.B., 532  
Adams, R.E., 550  
Adler, D., 274–275  
Adler, M.D., 68, 418, 752  
Ahmed, A.S., 548  
Akerlof, G., 494, 767  
Albin, E., 418  
Albiston, C., 252  
Alborn, T., 491  
Alchian, A., 270, 544, 546  
Aldy, J.E., 752  
Alexander, M., 246  
Alexander, V., 711  
Alicke, M.D., 147n8, 337, 340  
Allais, M., 143, 164, 346–347  
Allcott, H., 727, 730  
Allen, R.J., 280  
Allen, S.G., 106  
Allyn, J., 257  
Alm, J., 604  
Aloysius, J.A., 344  
Alschuler, A., 651, 660  
Alter, A., 118, 428–429, 431  
Alterman, R., 384n2  
Ambrus, A., 49  
Amir, O., 216n1, 220n5, 222, 227, 684  
Anastosopoulos, V., 531–532  
Anderson, C.J., 272, 630  
Anderson, C.M., 48  
Anderson, E., 775  
Anderson, J.C., 360, 690–691  
Anderson, J.R., 248–249  
Anderson, M.C., 680  
Andreoni, J., 37, 201  
Andrews, E.L., 518  
Angelova, V., 415–416, 419, 431  
Aquino, K., 216, 231  
Areed, P.E., 540, 553  
Ariely, D., 51, 216n1, 219–220, 222, 225, 227, 630, 684, 773  
Aristotle, 119  
Arkes, H.R., 15, 271, 273, 547, 629, 675, 684n7  
Arlen, J., 306, 316–317, 323, 426, 518–519, 539, 549  
Armor, D.A., 340  
Armstrong, M., 467, 471, 545, 555–556  
Aronson, E., 257, 399, 722  
Arrow, K., 447, 494–495, 505  
Asch, D.A., 80  
Asch, S., 257  
Asher, M.A., 623  
Attanasi, G., 415–416, 419, 431  
Aujila, P.S., 633  
Avineri, E., 347  
Aviram, A., 345, 539, 556  
Axelrod, R., 29  
Ayal, A., 542  
Ayal, S., 219, 224–225  
Ayres, I., 252, 291, 323, 379, 390–391, 481, 484, 758  
Ayton, P., 15  
Babcock, L., 138, 156, 216n2, 224, 231, 276, 278, 338, 626–628, 646–647  
Badea, C., 428n13, 429  
Badgio, P., 21, 76  
Bagenstos, S.R., 107, 422  
Bailey, E.M., 544, 554  
Bailey, M.A., 665n1  
Bailis, D.S., 627

- Bainbridge, S.M., 527–528  
Baker, J.A., 171  
Baker, L.A., 343, 396, 446, 452  
Baker, T., 342, 491–492, 499–500, 505, 509, 512, 580  
Bakos, Y., 275, 465, 472  
Balcetis, E., 217  
Baldassarri, D., 50  
Baldwin, J.R., 553  
Balliet, D., 35, 43, 48  
Banaji, M.R., 219, 220–221, 226, 230, 684  
Bandura, A., 223  
Banerjee, A.V., 155  
Barber, B.M., 340  
Barberis, N.C., 97, 498, 506  
Bargh, J.A., 213, 216, 683  
Bar-Gill, O., 274, 340–341, 470–473, 474n5, 477–481, 483, 485n21, 512, 542, 551, 601–602n2, 602, 723, 732  
Bar-Hillel, M., 272, 287, 305, 682–683, 735  
Barkan, R.S., 225  
Barker, R.A., 711  
Barnard, C., 98  
Barnett, R., 444  
Baron, J., 4–9, 14–16, 18–23, 28, 62, 66–68, 70–85, 172, 183–184, 272–273, 278, 286–287, 316, 355–356, 378, 411, 427, 460–461, 575, 601n1, 602, 608–611, 614, 618, 666, 674, 755  
Baron-Cohen, S., 35  
Barr, M.S., 484  
Barringer, F., 253  
Barro, R., 605  
Barros, D.B., 378–379, 382, 388  
Bartels, B., 258  
Bartels, D., 76  
Barzun, C.L., 707  
Bastardi, A., 633  
Batson, C.D., 223  
Baucells, M., 347–348  
Baum, L., 665  
Baumeister, R.F., 271  
Bayazit, M., 177  
Bayes, T., 3–4, 8, 11, 667, 669, 708–709, 711–712  
Bazerman, M.H., 214, 217, 219–221, 226–227, 231–232, 270, 356, 392, 632, 635  
Beach, L.R., 3, 356  
Beal, E.D., 496  
Beales, H., 469  
Beattie, J., 82  
Bebchuk, L., 208, 519, 550, 623, 625  
Becher, S., 275, 456n1, 486  
Beckenkamp, M., 552  
Becker, G., 113, 158–160, 169, 196, 245, 543, 569–570, 572, 575–576, 584n12, 600, 752  
Beecher-Monas, E., 708  
Beggan, J.K., 385, 394  
Belk, R.W., 385–386, 394  
Bell, D.E., 4, 278  
Bell, E., 607  
Benartzi, S., 22, 102, 179, 271, 606  
Benedict, K., 359n1  
Benkler, Y., 214  
Bennett, J., 74  
Bennett, M., 543, 552, 555, 559  
Ben-Shahar, O., 105, 279, 324, 345–346, 477, 479, 481, 485, 503  
Bentham, J., 68, 570n2, 571n4, 590–591, 595  
Benton, J., 217  
Bereby-Meyer, Y., 219, 232  
Berg, J., 33, 550  
Bernstein, A., 248, 523  
Bernstein, L., 203  
Bernstein, M.H., 368  
Berscheid, E., 63  
Bersoff, D.M., 220n5, 222  
Bersoff, D.N., 705  
Beshears, J.J., 737  
Bezdeka, R.H., 760  
Bhardwaj, V., 305–306, 393  
Bhole, B., 414  
Bibas, S., 643, 655  
Bicchieri, C., 65  
Biddle, G.C., 547  
Bigus, J., 413  
Bilz, K., 82, 85, 146n5, 225, 246, 255, 258, 740  
Birke, R., 626, 647  
Birnbaum, M.H., 9, 176, 271n1  
Bisin, A., 723  
Bjerk, D., 657  
Blader, S.L., 630  
Blair, M., 208, 518, 529  
Blanck, P.D., 664  
Blinder, A.S., 528  
Block-Lieb, S., 398  
Bloomer, C., 273  
Blumenthal, J.A., 102, 273, 323, 379, 398

- Blumer, C., 629, 675  
 Blumstein, J.F., 368  
 Blundell, R., 132  
 Board, O., 470, 480, 483  
 Boari, N., 645  
 Boehner, J., 615  
 Bohner, G., 253  
 Bohnet, I., 227  
 Bokhari, S., 273  
 Boldry, J.G., 428  
 Boles, T.L., 278  
 Bolger, F., 337  
 Bollinger, B., 728  
 Bolton, G.E., 35  
 Boonen, L., 272  
 Borgida, E., 259  
 Bornstein, B.H., 422, 676–681, 692  
 Bornstein, G., 49  
 Bossaerts, P., 344  
 Boswell, J., 339  
 Bower, D.J., 256  
 Bowles, S., 32, 35  
 Boyce, R.R., 315  
 Boyd, C.L., 623, 762  
 Boyd, R., 29  
 Bracha, A., 51, 499  
 Brainerd, C., 173  
 Braithwaite, J., 688  
 Braithwaite, V., 688  
 Braman, D., 246, 754  
 Brandeis, L., 588  
 Branscombe, N.R., 364  
 Branson, D.M., 530–532  
 Braucher, J., 138  
 Braun, M., 493, 498, 500, 506  
 Brawner, D., 254  
 Brehm, J.W., 248  
 Brekke, K.A., 37  
 Breneman, J., 137  
 Brennan, S., 286  
 Brenner, L., 525, 626, 634  
 Brewer, N., 666  
 Breyer, S., 23, 348  
 Brickman, D., 258  
 Brickman, L., 276  
 Brickman, P., 101, 422–423, 591, 593  
 Brighton, H., 177  
 Brinig, M.F., 133  
 Brockner, J., 273  
 Bröder, A., 10  
 Bromley, D.W., 770  
 Bronchetti, E.T., 737  
 Bronsteen, J., 101, 108, 577n6, 591, 593–594  
 Broome, L.L., 530  
 Brosnan, S.F., 314  
 Brown, D.A.H., 531–532  
 Brown, D.L., 531–532, 585  
 Brown, D.J., 499  
 Brown Jr., G.M., 302  
 Brown, J.P., 407  
 Brown, J.R., 740  
 Brown, T.C., 309  
 Brown, T.A., 259  
 Bruce, J., 501  
 Brudney, V., 519–520, 522, 529  
 Brunswik, E., 12, 173n5  
 Bubb, R., 102, 472, 478  
 Buccafusco, C.J., 101, 108, 126, 320–321, 379,  
     390, 577n6, 591, 593–594  
 Bucciarelli, M., 77  
 Buckholtz, J.W., 40  
 Budescu, D.V., 6, 13, 345  
 Budzinski, O., 139  
 Bui, T.M., 760  
 Bullens, L., 630  
 Burk, D.L., 361  
 Burke, R.J., 531  
 Burrows, L., 216, 683  
 Burrows, P., 756  
 Bush, G.W., 615–617  
 Bushey, B., 20–21  
 Byun, S.-E., 274  
 Cabantous, L., 344  
 Cacioppo, J.T., 755  
 Cai, H., 685  
 Cain, D., 527  
 Calabresi, G., 207, 327, 390, 405, 407, 410  
 Caldeira, G.A., 254, 258n5  
 Callen, C., 705  
 Camerer, C.F., 23, 115, 199, 271, 290, 340, 344,  
     384, 419, 478, 485, 493, 499, 503, 545, 556,  
     600, 633, 649, 724, 756



- Cameron, D., 720  
Cantril, H., 113, 147n17  
Caplan, R.A., 356  
Carasso, A., 607  
Carbado, D.W., 530  
Card, D., 723  
Cardi, W.J., 129  
Carlin, B.I., 106  
Carlson, A.E., 250–251  
Carlson, C.A., 138  
Carlson, K.A., 175  
Carpenter, J., 44  
Carroll, G.D., 735, 737  
Carter, D.A., 532  
Caruso, E.M., 219  
Carver, C.S., 340  
Cary, W.L., 522  
Casari, M., 47  
Casper, J.D., 359n1, 677  
Cassens Weiss, D., 243  
Chambers, J.R., 338, 340  
Chammah, A.M., 29  
Chamrand, T.L., 683  
Chan, C., 501  
Chandler, A., 528  
Chandler, C.C., 342  
Chandler, S., 497  
Chapman, G.B., 348, 678–679  
Chapman, J.P., 547  
Chapman, L.J., 547  
Charness, G., 33, 35–36  
Chartrand, T.L., 216  
Chase, O.G., 676, 681  
Chaudhuri, A., 34–35, 41, 43–44  
Chen, K.K., 668  
Chen, M.K., 109, 216, 314, 683  
Cheney, F.W., 356–357  
Chesson, H., 345  
Chetty, R., 602, 609, 611–613, 725, 739  
Chiteji, N., 733  
Choe, S.Y., 76  
Choi, J., 606  
Choi, S.J., 665  
Choi, S., 503, 527  
Chomsky, N., 65  
Christensen-Szalanski, J.J., 355, 359, 628  
Christner, J., 273  
Chuang, W., 525  
Chugh, D., 219, 226, 232  
Cialdini, R.B., 250, 252, 257, 274  
Citrin, J., 258  
Clark, G., 491  
Clark, R.L., 106  
Clark, S.E., 137  
Clarke, A., 256  
Clayton, C.W., 665n1  
Clemens, M.A., 256  
Clermont, K.M., 709  
Coase, R., 97, 300–302, 310–311, 313, 329, 405, 408, 519, 546  
Coates, D., 422–423, 591, 593  
Cochon, A., 533  
Cohen, A., 494, 519  
Cohen, D., 138, 304, 330, 389  
Colby, A., 62  
Colman, A.M., 345  
Comanor, W.S., 558  
Condorcet, Marquis de, 155–156  
Conly, S., 103  
Coon, D., 680  
Cooper, A.C., 273  
Cooper, J., 248, 399, 558  
Cooter, R., 37, 97, 108, 111, 225, 247, 395, 410, 624–625, 627, 631  
Corbin, A., 441  
Correll, J., 684  
Correll, S., 252  
Costa, D.L., 758  
Costa Dias, M., 132  
Costanza, R., 771  
Costanzo, M.A., 722  
Coursey, D.L., 303  
Covey, R., 341, 649  
Cox, M., 671, 673  
Crandall, C.S., 272  
Crane, D.A., 418  
Crano, W.D., 259  
Cranston, M.W., 283  
Craswell, R., 390, 469, 479n11  
Crisp, R., 590  
Croson, R., 44  
Cross, F.B., 283  
Csikszentmihalyi, M., 386  
Cubitt, R., 35, 38–39

- Cullison, A.D., 708  
 Cupples, J., 259  
 Currie, D., 284  
 Currie, J.S., 735  
 Cushman, F., 73, 219  
 Cutler, B.L., 664  
 Cutler, D., 496–498, 511
- Daftary-Kapur, T., 672  
 Dagan, H., 380  
 Dahrendorf, R., 767  
 Dana, D., 760  
 Dana, J., 6, 13, 222  
 Dan-Cohen, M., 159n33  
 Daniels, K.C., 343  
 Darley, J.M., 82, 85, 109–110, 245–246, 255, 257,  
 428–429, 431, 584, 629, 635, 667  
 Daston, L., 491  
 Dau-Schmidt, K., 203  
 Davidai, S., 252  
 Davidson, D., 3  
 Davis, K., 473  
 Davis-Stober, C.P., 6, 13  
 Dawes, R., 13, 167n1, 198, 446  
 Dawson, N.V., 364–365  
 Dayan, E., 735  
 De Cremer, D., 220n5  
 DellaVigna, S., 115, 271, 480, 524, 526  
 Demsetz, H., 197, 544, 546  
 Denrell, J., 173  
 Depoorter, B., 184, 627  
 de Quervai, D.J.F., 632  
 Desai, S.D., 228  
 DeScioli, P., 273  
 Devine, D.J., 154n23, 690  
 Devine, P.G., 232  
 Dharmapala, D., 257  
 Dhir, A., 532  
 Dholakia, R., 254  
 Diamond, S., 378–379, 383, 676–677, 679  
 Dias, M.G., 63  
 Dickel, N., 253  
 Dickerson, A.M., 526  
 Dickhaut, J., 33, 550  
 Diener, C., 422  
 Diener, E., 101, 422
- DiPerna, P., 760  
 Dittmann, I., 271  
 Dixon, H.D., 555  
 Doane, M.J., 303  
 Dodge, M., 671  
 Dogan, E., 338  
 Dohmen, T., 501, 503n3  
 Doll, Jen, 711  
 Donkers, B., 272  
 Donnelly, J., 282  
 Drolet, A., 316, 393  
 Drouvelis, M., 35, 38–39  
 Du, N., 345  
 Dubourg, W.R., 304  
 Duellman, S., 548  
 Duff, A., 570–571  
 Duflo, E., 723  
 Dufwenberg, M., 34–37, 39  
 Dumas, R., 672  
 Dunn, E., 102  
 Dunne, T., 553  
 Dunning, D., 217, 225, 337–339, 342  
 Dunwoody, P.T., 173n5  
 Durkheim, E., 594–595  
 Durose, M.R., 712–713  
 Dworsky, M., 733  
 Dysart, J.E., 138
- Easterbrook, F.H., 518, 520–523, 557–558  
 Easterlin, R.A., 769  
 Eberhardt, J.L., 684  
 Ebert, J.E.J., 630  
 Eckel, C., 37  
 Edgeworth, F.Y., 196  
 Edlin, A., 554  
 Edwards, W., 3, 167n1, 178–179, 180n10  
 Edward VII (king of England), 453  
 Ees, H.V., 524  
 Effron, D.A., 217n3, 223  
 Egan, P.J., 258  
 Egas, M., 48  
 Ehrhart, K.-M., 45  
 Eichenberger, R., 64  
 Eidelman, S., 272  
 Eidenmüller, H., 485, 486n22, 769, 773  
 Eigen, Z., 457n2

- Einhorn, H.J., 345  
Eisenberg, M., 206, 341, 444, 450, 460, 462  
Eisenberg, T., 129, 623, 666, 671, 677, 679,  
    690n9, 706  
Eldar, O., 219, 232  
Ellickson, R., 31, 203, 225, 384–385, 389  
Ellinghaus, M.P., 687, 692  
Ellison, G., 466n1, 475  
Ellison, S.F., 475  
Ellsberg, D., 344  
Ellsworth, P., 113, 255  
Elstad, B., 531  
Elster, J., 146, 605, 618, 767, 773  
Emery, R.E., 343, 396, 446, 452  
Emmanuel, A., 253  
Engel, C., 30, 32, 44, 50, 115, 125–126, 129, 135,  
    170, 177–178, 292, 546, 560–561, 574, 646,  
    709, 729  
Engelmann, D., 36, 322  
Engen, R., 653  
Englich, B., 678–679  
Ensminger, J., 53  
Epley, N., 219, 225, 678  
Epstein, L., 665n1  
Epstein, R., 206–207, 282  
Epstein, S., 18  
Erev, I., 273, 362  
Erickson, M.A., 137  
Ericson, K.M.M., 304  
Ericsson, K.A., 6  
Erkut, S., 531  
Erskine, H., 244  
Ert, E., 273  
Etzioni, A., 391  
Evans, J., 18, 215  
Evans, L.B., 553  
  
Fagan, J., 247  
Fairfax, L.M., 530, 532–533  
Falk, A., 30, 35–37  
Falkinger, J., 50  
Fancsali, B.L., 431  
Faraday, M., 20–21  
Farber, H.S., 632, 771  
Farley, J., 770–771  
Farley, L., 243  
  
Farnsworth, E.A., 369  
Farnsworth, W., 320, 392n9  
Farrell, K.A., 532  
Faure, M.G., 410  
Fazio, R.H., 248  
Feeley, M., 659  
Fehr, E., 30, 33, 35–40, 42, 46–47, 203, 632  
Fein, S., 672  
Feldbrugge, F.J.M., 71  
Feldman, M., 225  
Feldman, Y., 82, 221, 224–226, 230, 346, 363,  
    418, 457n2, 460, 605n4, 687  
Feldstein, M., 447, 498  
Felstiner, W.L.F., 623  
Fennell, L., 116n20, 378  
Fernandes, D., 105  
Ferrari, F., 479, 481  
Ferraro, P.J., 339  
Ferreira, D., 532  
Ferrell, A., 519  
Ferris, S.P., 525, 548  
Festinger, L., 257, 399, 765, 768  
Fetherstonhaugh, D., 761  
Feygina, I., 766  
Fiedler, K., 173  
Fiedler, P., 243  
Finkelstein, A., 739  
Finkelstein, S.R., 272  
Finucane, M., 754  
Fiorentini, G., 645  
Fisch, J., 525  
Fischbacher, U., 30, 35–37, 39–40, 42–44  
Fischel, D., 518, 520–523, 557  
Fischel, W.A., 385  
Fischer, G.W., 289  
Fischhoff, B., 21, 148, 180–181, 354–355, 364,  
    410, 673, 710, 761  
Fisher, G., 651, 657  
Fiske, A.P., 70, 772  
Fiske, S.T., 338  
Fisman, R., 85  
Flavell, J.H., 63  
Fletcher, G., 570  
Fobian, C., 628  
Foley, R.J., 243  
Fong, G.T., 19  
Foot, P., 71

- Forbes, D.P., 528  
 Förster, J., 630  
 Forsythe, R., 32  
 Fox, C.R., 345, 547, 626  
 Frakes, M., 132  
 Francis, J., 456  
 Francis, K.E., 341  
 Frank, J., 664  
 Frank, R.H., 107, 110, 769  
 Frankfurt, H.G., 771  
 Franklin, C.H., 258  
 Frederick, S., 10, 17–19, 171, 277, 316, 422, 724  
 Fredrick, S., 501, 503n7  
 Freedman, J.L., 257  
 Freeling, A.N.S., 4  
 Freeman, D., 216  
 Frey, B., 44, 209, 604, 773  
 Fried, J., 208, 550  
 Friedman, D., 32  
 Friedman, M., 443, 523, 542  
 Friedman, R.D., 280  
 Friedman, R.D., 713  
 Frisch, D., 19  
 Froeb, L.M., 540–541, 543, 557, 705  
 Fryer, R.G., 271  
 Fuller, L., 444  
 Funder, D.C., 170  
 Funk, C.L., 257–259  
 Furedi, F., 349  
 Fuster, A., 304
- Gabaix, X., 470, 475, 507, 542  
 Gabrielsson, J., 524  
 Gächter, S., 30, 34–35, 37–39, 41–53, 64, 198, 415n7, 632, 666  
 Gaertner, L., 428  
 Gaeth, G.J., 722  
 Gal, D., 272, 274  
 Galanter, M., 461, 623–624  
 Galbiati, R., 37  
 Gale, W., 733, 735  
 Galin, A., 273  
 Galinsky, A.D., 225, 627  
 Gallagher, K.M., 179n9  
 Galle, B., 599, 609, 611, 612n7  
 Gamage, D., 599, 602, 609–613
- Ganzach, Y., 16  
 Garcia, S.M., 556  
 Gardbaum, S., 283  
 Gardner, R., 46  
 Garnett, N.S., 133  
 Garoupa, N., 230, 568n1, 577n6, 585–586  
 Gaskins, I.W., 21, 76  
 Gautschi, T., 201  
 Gavison, R., 283  
 Gay, S., 280  
 Gayer, T., 762  
 Gazal-Ayal, O., 132, 649  
 Gelman, S.A., 22  
 Geltner, D., 273  
 Gely, R., 675  
 Georgantzis, N., 303, 318  
 George, T.E., 280, 674  
 Gerla, H.S., 539, 554  
 Geroski, P.A., 553  
 Gerrard, M., 342  
 Gerritsen, L.E.M., 344  
 Gertner, R., 323  
 Gervais, S., 525  
 Ghosh, S., 415  
 Gibbons, F.X., 342  
 Gibson, J.L., 254, 258n5  
 Gifford, G., 432  
 Gigerenzer, G., 12, 19, 172–173, 177, 181, 645, 761  
 Gilbert, D.T., 101–102, 257, 271, 630  
 Gillette, C., 469, 482  
 Gillman, H., 665n1  
 Gilo, D., 410n3, 475  
 Gilovich, T., 167n1, 252, 316, 629, 678  
 Gilson, R.J., 634–635  
 Gino, F., 214, 216–219, 224–225, 227–228, 232, 356  
 Ginsburg, D.H., 551, 559  
 Ginsburg, T., 320  
 Gintis, H., 28, 30, 32, 35  
 Girotto, V., 18  
 Giuliani, R., 588  
 Glaeser, E.L., 290, 469  
 Glaser, M., 548  
 Glimcher, P.W., 32  
 Glöckner, A., 10, 646  
 Gneezy, U., 37, 204, 378, 545, 587n15  
 Godsil, R.D., 379  
 Goetz, C.J., 279, 448

- Goff, P.A., 171  
Golann, D., 632  
Gold, V.J., 706, 708, 711, 713  
Goldman, A.I., 714  
Goldsmith, R.E., 766  
Goldstein, D., 12, 99–100, 272, 483  
Goldstein, N., 257  
Goldstein, W.M., 173n5  
Gonzales, M.H., 722  
Goode, M.R., 216  
Goodman, J.C., 284  
Goodman-Delahunty, J., 341, 626  
Goodsell, C.A., 138  
Gordon, W.J., 282  
Govorun, O., 147n8, 337  
Gowda, R., 82  
Graff, W., 251  
Graham, J.D., 751  
Graham, S., 684–685  
Grechenig, C., 49  
Green, J.D., 339  
Greene, E., 671, 676–681, 692  
Greene, J.D., 18, 65–66, 68, 72–73, 75–76,  
83–84, 219–221  
Greenfeld, L.A., 712  
Greenfield, K., 151n18, 209, 341, 518, 520,  
527–530, 533  
Greenspan, A., 518  
Greenstone, M., 729, 733  
Greenwald, A.G., 220n5, 684nn6–7, 765  
Gregg, A.P., 340  
Greiner, B., 49  
Griffin, D., 167n1, 525  
Griskevicius, V., 257  
Grogger, J., 576  
Gronlund, S.D., 138  
Gross, S.R., 255, 625, 628  
Grosskopf, O., 282  
Grossman, G., 50  
Grossman, P., 37  
Grossman, S.J., 480  
Grubb, M.D., 480  
Gruber, J., 605, 739  
Grujic, J., 29, 41  
Grushka-Cockayne, Y., 12  
Guilbault, R.L., 355, 359, 364, 673  
Gulati, M., 530, 665  
Gürçay, B., 76  
Gürerk, Ö., 47  
Gustafson, J.L., 531  
Güth, W., 32  
Guthrie, C., 167, 175, 276–280, 326, 358–359,  
580, 623, 627–632, 634–636, 670, 672–674,  
678, 680, 683, 690–691, 693, 706–707,  
710, 714  
Guttel, E., 155n27, 410n3, 419n10, 420, 581,  
605n4, 672n3, 704, 706, 709  
Guyatt, V., 259  
  
Habermas, J., 771  
Haddad, B.M., 740  
Haidt, J., 39, 62–63, 77, 183, 219  
Haigh, M.S., 114, 273  
Haigner, S., 50  
Haisley, E., 231, 756  
Halbersberg, Y., 155n27, 430  
Halevy, Y., 345  
Hall, K., 526, 530  
Hall, M., 133, 138  
Hamann, H., 526  
Hamilton, C., 592  
Hamilton, J.T., 419, 732  
Hammack, J., 302  
Hand, L., 406, 409, 411  
Handel, B.R., 512  
Haneman, V., 456  
Hanemann, W.M., 393n11  
Hannaford-Agor, P., 624  
Hans, V.P., 174, 666, 671, 675, 677–679, 692  
Hanson, J.D., 167, 275, 767–768  
Harbaugh, W.T., 306  
Harcourt, B.E., 588  
Hare, R.M., 67–69  
Harel, A., 182, 230, 419n10, 420, 423n12, 496,  
568n1, 570, 574, 577, 579–581, 593n19, 594,  
605n4, 687  
Harinck, F., 273  
Harless, D.W., 306  
Harnay, S., 586n14  
Harris, R., 81, 418  
Harris, S., 339  
Harrison, D.A., 214  
Hart, A.J., 676, 680–681

- Hart, O.D., 480  
 Hartman, R.S., 303  
 Harvey, O.J., 256  
 Harvey, S., 217  
 Hassin, R.R., 684  
 Hastie, R., 170, 355, 359–360, 412n4, 424–425, 632, 667–669, 673, 676, 678–679, 691, 706  
 Hastorf, A., 113  
 Hatfield, E., 755  
 Haubl, G., 317–318  
 Hauser, J., 15  
 Hauser, M., 73, 219  
 Hawkins, S.A., 355  
 Hayden, B.Y., 344  
 Healy, P.J., 626  
 Heath, C., 273, 337–339, 342, 548, 581, 722, 730, 754  
 Heath, D., 722, 730  
 Heaton, J.B., 525  
 Heckman, J., 30  
 Heffetz, O., 304  
 Hegel, G.W.F., 382  
 Heimer, C., 495  
 Heine, S.J., 51  
 Heinzerling, L., 750, 761  
 Heise, M., 627  
 Heldman, C., 258  
 Hell, W., 364  
 Helland, E., 635  
 Heller, K.J., 348, 707  
 Helstad, O., 456  
 Helzer, E.G., 339  
 Hendel, I., 495  
 Hennig-Schmidt, H., 34, 37, 39  
 Henrich, J., 32–33, 40, 50–51, 53, 200, 632  
 Hensler, D.R., 623  
 Hermalin, B.E., 550  
 Herrmann, B., 35, 41, 43, 48, 51–53  
 Hersch, J., 690n9  
 Hersch, P.L., 532  
 Hershey, J.C., 355–356, 411  
 Hertwig, R., 76, 135, 173n4, 362  
 Heukamp, F.H., 347–348  
 Heumann, M., 655  
 Hewstone, M., 429  
 Heyman, D., 586, 773  
 Hibbing, J.R., 258n5  
 Hilbert, M., 173  
 Hill, C.A., 529, 599, 771  
 Hill, G.W., 550  
 Hill, S., 762  
 Hillman, R.A., 154, 155n26, 164, 465  
 Hills, A., 586  
 Hines, J., 602, 614  
 Hinloopen, J., 560  
 Hinsz, V.B., 550, 680  
 Hiriart, Y., 415–416, 419, 431  
 Hirsch, A.J., 182n11  
 Hirschhoff, J.T., 410  
 Hirshleifer, D., 723  
 Ho, J.L.Y., 345  
 Ho, S.S., 630  
 Hobbes, T., 28, 41  
 Hoeger, R., 219  
 Hoekstra, V.J., 258  
 Hoffman, D.A., 246, 460–461, 527–528, 623  
 Hoffman, E., 202, 391, 393, 396, 600  
 Hogarth, R.M., 13, 175, 274, 345, 499, 545, 628  
 Hollander, E., 254, 630  
 Hollander-Blumoff, R., 630  
 Hollard, G., 322  
 Hollis, J., 220n5  
 Holmes, O.W., 195–196, 439, 457, 459  
 Holmes, S., 283  
 Holt, John, 439  
 Holyoak, K.J., 667–669  
 Holzberg, A.D., 338  
 Hoogendoorn, S., 532  
 Hoover, S., 133  
 Horowitz, J.K., 270, 303, 305  
 Horton, T.J., 539, 557  
 Houston, D.A., 256  
 Hovenkamp, H., 319, 540, 553  
 Hovland, C.I., 254, 256  
 Howarth, R.B., 740  
 Hrebec, D., 632  
 Hsee, C.K., 231, 347, 499–500, 724, 754  
 Huang, P.H., 527  
 Hubbell, B., 752  
 Huber, J., 634  
 Huber, V.L., 392  
 Huberman, G., 738  
 Huck, S., 348, 545, 555–556  
 Huffman, M., 540, 551

- Huigens, K., 182  
 Hume, D., 31, 120  
 Hunt, C.V., 259  
 Huo, Y.J., 246–247  
 Hurley, N.L., 254  
 Huse, M., 524, 531  
 Hutchins, R.M., 703  
 Huysentruyt, M., 505
- Idson, L.C., 18  
 Indhal, K.E., 680  
 Ingley, C., 531  
 Inman, J.J., 274  
 Irlenbusch, B., 47  
 Isenberg, D.J., 530, 689  
 Isoni, A., 308–309, 384n3, 601n2  
 Issacharoff, S., 156, 396, 419  
 Iwry, J., 733, 735  
 Iyengar, S., 739
- Jackson, H., 502  
 Jacob, K., 751  
 Jacobson, P.D., 248–249  
 Jaffee, D., 494  
 Jami, A., 216  
 Janger, E.J., 398  
 Janis, I.L., 530  
 Janoff-Bulman, R., 422–423, 591, 593  
 Jayaraman, N., 525, 548  
 Jegen, R., 209  
 Jensen, M., 521, 546  
 Jiang, W., 738  
 Jimenez, D., 105  
 Jindal, P., 499, 505–506  
 Johnson, E.J., 98–100, 272–273, 317–318, 483, 484n15, 497, 500, 506, 510, 678–679  
 Johnson, J.E., 501  
 Johnson, N.D., 34  
 Johnson, S.R., 280  
 Johnson, T.R., 258  
 Johnson-Laird, P.N., 18, 77  
 Johnston, J.S., 182n11  
 Joireman, J., 43  
 Jois, G.U., 674  
 Jolls, C., xin1, 115–117, 150, 151n18, 156–157, 162–163, 167, 170–171, 175, 213, 228, 337, 340, 362, 366–367, 379, 390–391, 392n9, 396, 398, 412, 518, 530, 533, 539, 568n1, 571, 584, 599, 648, 706, 709, 724, 752, 764
- Jones, E.E., 429n14  
 Jones, O.D., 314  
 Jones, O., 117, 183  
 Jones, S., 722  
 Jones-Lee, W., 304  
 Joo, T.W., 532  
 Joseph, C., 62, 219  
 Jost, J.T., 765–766  
 Jourden, F., 177, 280, 327–328, 391–392, 670  
 Joyce, E.J., 547  
 Joyce, M.A., 81  
 Judd, C.M., 428  
 Juslin, P., 173  
 Just, D., 735
- Kabalin, R., 38  
 Kagan, S., 286, 288, 751  
 Kagehiro, D.K., 709  
 Kagel, J.H., 391  
 Kahan, D., 31, 220, 225, 231, 246, 248, 251–253, 256–257, 587–588, 710–711, 713, 754–755  
 Kahane, G., 78n6  
 Kahn, M.E., 758  
 Kahneman, D., 3, 9, 11, 17–18, 20, 28, 64, 79, 82, 96, 102, 110, 112, 126, 143, 148, 149nn12–13, 154–155, 163, 167–169, 171–175, 178–181, 213, 215, 224, 268, 270–273, 277, 302, 305–309, 312, 328, 347, 384, 388, 391, 393, 394n12, 413, 421, 424–425, 445–447, 466n2, 485n20, 500, 506, 525, 545, 547, 549, 554, 559, 573, 575, 584, 590–591, 593, 601n1, 609, 627–628, 636, 647, 674, 676–678, 681–682, 689–690, 703–705, 722–724, 752–753, 757, 760, 766
- Kameda, T., 550  
 Kamenica, E.S., 481, 739  
 Kamin, K.A., 356, 365, 411, 673  
 Kamm, F.M., 286  
 Kang, J.M., 686  
 Kant, I., 78, 570n3  
 Kanter, R.M., 531  
 Kaplan, J., 708  
 Kaplow, L., 182n11, 390, 395, 610–611, 613, 713  
 Karelaia, N., 13



- Karla, A., 499  
 Karras, M., 687, 692  
 Kassin, S.M., 156  
 Katyal, S., 255, 391  
 Katz, A.W., 634  
 Kauder, N.B., 244  
 Kawakami, K., 216  
 Kay, A.C., 216  
 Kaye, D., 705  
 Keefer, P., 210  
 Keeton, R.E., 426  
 Keeton, W.P., 426  
 Keinan, A., 317–318  
 Keller, C., 500  
 Keller, L.R., 344  
 Kelly, J.J., 379  
 Kelly, J.R., 359n1  
 Kelman, H.C., 399  
 Kelman, M., 115, 172, 292, 313–314, 634, 670  
 Kempton, W., 22  
 Kennedy, D., 301, 397n13, 686  
 Kennedy, K.A., 632  
 Keren, G., 344, 760  
 Kermer, D.A., 271  
 Kern, M.C., 232  
 Kerr, N.L., 550, 682, 688–689  
 Kerwin, J., 672  
 Keser, C., 45  
 Kesibir, S., 183  
 Kessler, E., 705  
 Khemlani, S., 77  
 Kidd, R.F., 707, 711, 713  
 Kidder, R.M., 220n5  
 Kim, A., 259  
 Kinder, D.R., 258  
 King, N., 652  
 King, R., 583n11  
 Kingsley, D.C., 309  
 Kinnaman, T.C., 249  
 Kipperberg, G., 37  
 Kirchgässner, G., 31  
 Kirchner, E., 37  
 Kirchsteiger, G., 33, 35–36  
 Kirk, D., 589  
 Kiser, R., 623, 629  
 Kish-Gephart, J.J., 214  
 Kiviniemi, M.T., 342  
 Klaes, M., 35  
 Klass, A.B., 368  
 Klass, G., 312  
 Klayman, J., 548  
 Klein, B., 558  
 Klein, G., 175, 547, 690  
 Klein, W.M., 162, 338, 342  
 Kleinbard, E.D., 290  
 Klick, J., 111, 118–119, 132, 175, 290  
 Kluegel, J.R., 396  
 Knack, S., 210  
 Knetsch, J.L., 64, 149n13, 224, 272–273, 277, 302, 304–310, 312, 384, 388–389, 391, 393, 549, 554, 600, 601n2, 647, 674, 757  
 Knight, F., 771–772  
 Knight, J., 665  
 Knight, J.H., 242–243  
 Knobe, J., 65  
 Knudsen, T., 78  
 Kobayashi, B.H., 705  
 Kocher, M.G., 43, 50  
 Kochi, I., 752  
 Koehler, D.J., 609, 626  
 Koehler, J.J., 19, 184, 280, 674, 708  
 Kogut, T., 79  
 Koh, W., 601n2  
 Kohlberg, L., 61–63  
 Kolber, A., 591–592, 593n19  
 Koller, S.H., 63  
 Konrad, A.M., 531  
 Konrad, K.A., 604  
 Kopel, D., 244  
 Koriat, A., 21  
 Kornhauser, L., 414–416, 418, 627, 643n1  
 Korobkin, R.B., xi, 101, 114–115, 149n11, 167–168, 170, 177, 183, 213, 269–270, 273, 277–280, 306–307, 319, 322–326, 366, 384, 389, 391, 393, 410, 412, 457n2, 465, 472, 478n7, 483, 518–519, 523–524, 527–528, 549, 600, 601n2, 606, 627, 629, 631–632, 634–637, 648–649, 690, 709, 711, 713, 721, 723, 733, 736, 739, 753–754, 757, 767, 774  
 Kosaki, L.C., 258  
 Kosfeld, M., 44  
 Kostant, P.C., 209  
 Köszegi, B., 270, 304, 549, 605  
 Kouchaki, M., 216

- Kovera, M.B., 664  
Kraakman, R.H., 426  
Krambeck, H.J., 51  
Kramer, G.P., 550, 688, 752  
Kramer, K., 531  
Kranton, R.E., 767  
Krantz, D.H., 16, 19, 175–176, 499–500  
Krause, K., 306  
Krauss, D.A., 673  
Krawiec, K.D., 530  
Kreuter, M.W., 342  
Krieger, L.H., 229  
Krier, J.E., 390, 419, 599–600, 760  
Kristel, O.V., 396  
Kritzer, H.M., 276, 635, 665n1  
Kroft, K., 602, 609, 611, 613, 739  
Krueger, J.I., 170  
Ku, G., 273  
Kuang, J.X., 222  
Kuenzler, A., 228n8, 767, 771, 773, 775  
Kugler, T., 580  
Kuhn, P., 33  
Kuhn, T., 111–112  
Kuhse, H., 74  
Kuklinski, J.H., 254  
Kunda, Z., 217  
Kundu, R.P., 415  
Kunen, J., 652  
Kunreuther, H., 82, 175–176, 344, 447,  
497–500, 502, 511  
Kuran, T., 248, 255, 759  
Kurschilgen, M., 135  
Kurzban, R., 183, 273  
Kysar, D.A., 167, 228n8, 275, 750, 754, 758,  
761–762, 767–771
- Labine, G., 357, 411  
Labine, S.J., 357, 411  
Lacetera, N., 250  
Ladegard, G., 531  
Lahav, A., 432n18  
Laibson, D., 470, 475, 507, 542  
Lakshminaryanan, V., 314  
Lamm, H., 689  
Landeo, C.M., 138  
Landes, W.M., 415, 553, 586, 665
- Landsman, S., 672, 676–677, 693  
Lane, K.A., 529, 768  
Langan, P.A., 712–713  
Langbein, J.H., 623  
Langbein, L., 244  
Langdell, C., 438–439  
Langer, J., 491  
Langer, M., 644  
Langevoort, D.C., 151, 156, 167–168, 213, 274,  
518, 523–527, 529, 532, 539  
Lanvers, C., 623  
Larouche, P., 560  
Larrick, R.P., 5, 278, 548  
Lau, R.R., 258–259  
Lawless, R.M., 138  
Lawsy, S.B., 346  
Lax, D.A., 631  
Laycock, D., 460  
Lazarus, R., 750  
Leamer, E.E., 132  
Lee, B., 525  
Lee, D., 108  
Lee, D.H., 673  
Lee, J., 147n8  
Lee, M.R., 684  
Lee, S.M., 668  
Leeman, R.F., 62  
Lefebvre-Pinard, M., 72  
Legrenzi, P., 18  
Leibenstein, H., 544  
Leiter, B., 117n21  
Le Mens, G., 173  
Lemley, M.A., 361, 390  
Lempert, R., 412n4  
Leo, R.A., 703  
Lepper, M.R., 364, 633, 636, 765  
Lerner, J.S., 70, 231  
Lerner, M.J., 63  
Leshner, S., 16, 70  
Leslie, C.R., 539, 554–555  
Leslie, P., 728  
Lessig, L., 241  
Leung, K., 630  
Levenson, J., 713  
Levenstein, M.C., 555  
Leventhal, H., 256, 722  
Levin, I.P., 722

- Levine, D.K., 110  
 Levinson, H., 526  
 Levinson, J.D., 175, 684–685  
 Levitin, A., 107  
 Levitt, S.D., 108, 114  
 Levmore, S., 282  
 Lewinsohn-Zamir, D., 42, 288, 323, 377–378, 382,  
     384, 386–387, 392–398, 459–462, 768, 775  
 Liberman, N., 348, 394n12  
 Lichtenstein, S., 21  
 Lieberman, J.D., 666, 673  
 Liebman, J., 601, 608  
 Liersch, M.J., 272  
 Lifshitz, S., 346  
 Liljenquist, K., 217  
 Limburg, K.E., 771  
 Lin, D.Y., 152, 152n20  
 Lincoln, A., 601  
 Lind, E.A., 630  
 List, J.A., 113–114, 273, 304, 306  
 Listokin, Y., 136  
 Liu, H.H., 345  
 Livermore, M.A., 751, 759, 763  
 Lizzeri, A., 495  
 Llewellyn, K., 281, 441, 686  
 Lloyd, F.J., 173  
 Lobel, O., 227  
 Loewenstein, G., 10, 18, 79–80, 156, 176, 216n2,  
     218–219, 220n5, 224, 231, 270, 274–275,  
     277, 338, 396, 422, 474, 527, 577n6, 605,  
     626, 633, 646–647, 754, 756, 760–761, 772  
 Loftus, E.F., 356, 626  
 Logue, K.D., 230  
 Loomes, G., 278, 304, 308–309, 315, 384n3, 500,  
     601n2  
 Looney, A., 602, 609, 611, 613, 739  
 López-Pérez, R., 37  
 Lord, C.G., 364, 633, 636, 765  
 Lott, J., 577n7  
 Lotwis, M.A., 244  
 Lovallo, D., 340, 559, 636  
 Lowe, D.J., 365, 690  
 Lowery, B.S., 684–685  
 Ludwig, J., 588  
 Luini, L., 47  
 Luna, E., 644  
 Lund, F.H., 147n7  
 Luppi, B., 152, 156n28, 156n32, 398, 504, 606n5  
 Lusardi, A., 104–105  
 Luttmer, E., 259  
 Lynch, J.G., 481  
 Lyon, T.D., 708  
 Maccoby, N., 257  
 MacCoun, R.J., 113, 550, 627, 629–631, 635, 667,  
     680, 688–689  
 MacDonald, D.A., 480  
 Macey, J.R., 522  
 Macis, M., 250  
 Maclay, K., 253  
 MacLeod, W.B., 426  
 Macnab, B., 305–306, 393  
 Macneil, I.R., 206  
 Madrian, B.C., 272, 483  
 Magat, W., 445–446  
 Mahan, L., 711  
 Mahar, H., 343  
 Maier-Rigaud, F.P., 552  
 Maitland, F.W., 440  
 Malmendier, U., 480, 526, 548, 557, 560  
 Maltzman, F., 665n1  
 Mandel, D.R., 393  
 Mandel, G., 358, 673  
 Maniloff, H., 255  
 Mansbridge, J.J., 391  
 March, J.G., 559  
 Marciano, A., 586n14  
 Marewski, J.N., 181  
 Margolis, H., 391  
 Markowitz, H., 270  
 Marks, S., 625, 631  
 Marnet, O., 518, 533  
 Marois, R., 40  
 Marotta-Wurgler, F., 275, 465, 472–473, 479  
 Marquardt, N., 219  
 Marshall, J.D., 306–307  
 Marti, M.W., 679  
 Martin, A.D., 258  
 Martinek, W.L., 665  
 Martinsson, P., 43  
 Masser, B., 256  
 Masur, J.S., 101, 108, 320, 577n6, 591, 593–594  
 Maug, E., 271

- Mayhew, P., 690  
Mazar, N., 216n1, 220n5, 222, 684  
Mazor, K.M., 631  
McAdams, R., 247, 257–258, 320, 568n1  
McAlister, L., 274  
McAuliff, B.D., 679  
McCabe, K., 33, 202, 391, 550  
McCaffery, E.J., 18, 20, 83, 155, 328, 385, 421,  
599–600, 601n1, 602, 604, 606–611, 614, 618  
McCaffery, J., 23  
McCarthy, A.M., 273  
McCauliff, C.M.A., 711  
McClennen, E.F., 387n6  
McCloskey, A.L., 672  
McClure, S., 474  
McConnell, K.E., 270, 303, 305  
McConnell, M., 617  
McCormick, C.T., 280  
McCoy, C., 652  
McCrary, J., 108  
McDaniel, P.R., 290  
McDonnell, B.H., 529  
McFadden, D.L., 168–169  
McFerran, B., 723  
McGarity, T.O., 750, 763  
McGorty, E.K., 422  
McGraw, A.P., 272, 396  
McGuire, W.J., 257  
McHughen, A., 755  
McKendry, J., 735  
McKenzie, C.R.M., 182n11, 272, 668  
McKibbin, W.J., 770  
McMorrow, S., 497–498, 502, 511  
McNaughton, D., 74  
McNeil, B.F., 431  
McNeil, B.J., 547  
McShane, B.B., 623  
Mead, N., 216  
Meadow, W., 708  
Meckling, W.H., 521, 546  
Medina, B., 103, 125, 283, 286, 291, 771  
Medvec, V.H., 316, 629  
Mehra, R., 271  
Meier, S., 44, 51  
Melamed, D., 207, 327, 390  
Mellers, B.A., 173n4, 272, 278–279  
Mendez, M.A., 703  
Merges, R.P., 390  
Merkel, D., 593  
Merrill, T.W., 380, 392, 766n1  
Merritt, A.C., 217n3, 223  
Merry, S.E., 623  
Messé, L.A., 682  
Messick, D.M., 223  
Meyer, D., 256  
Meyerowitz, J.A., 338  
Meyerson, M., 456  
Mezulis, A.H., 338, 340  
Michaud, P.-C., 105  
Michelman, F.I., 377, 381  
Miguel, E., 85  
Mikhail, J., 62, 78  
Mikos, R., 576  
Mikulincer, M., 219  
Milgram, S., 201, 427, 529  
Milgrom, P., 480  
Milinski, M., 50–51  
Mill, J.S., 4, 13–14, 196  
Miller, D.T., 63, 223, 257, 259, 272, 632  
Miller, G.P., 634  
Miller, J.G., 74  
Miller, J.H., 201  
Miller, R.E., 623  
Milliken, F.J., 528  
Millimet, D.L., 113  
Millon, D., 325  
Milosavljevic, M., 10  
Min, K., 76  
Minsk, E., 66, 72, 674  
Mirabile, R.R., 255  
Mirrlees, J.A., 599  
Mislin, A.A., 34  
Mitchell, G., 111, 118–119, 171, 175–181, 182n11,  
183, 290, 504, 539, 544, 561, 666, 714, 725,  
756, 771  
Mitchell, O.S., 104–105  
Mitzkewitz, M., 140  
Mnookin, R.H., 625, 627, 631, 634–635, 643n1  
Modigliani, F., 605  
Moeller, S.B., 557  
Molot, J.T., 419  
Monin, B., 217n3, 223  
Moore, D.A., 218–219, 220n5, 224, 338, 340,  
527, 626, 633, 635

- Moore, D.W., 551, 559  
 Moran-Ellis, Jo, 256  
 Morck, R., 529  
 Moreno, J.A., 714  
 Moretto, G., 76  
 Morewedge, C.K., 149n13, 385n4, 388  
 Morgan, J., 528  
 Morgan, M.G., 762  
 Morgenstern, R.D., 760  
 Moro, A., 723  
 Morrill, M.S., 106  
 Moshinsky, A., 272, 287  
 Moshirnia, A., 687  
 Moskowitz, G.B., 684  
 Mueller, D.C., 557  
 Mueller-Langer, F., 410  
 Muermann, A., 493, 500, 506  
 Mulder, L.B., 35, 48  
 Mullainathan, S., 481, 484  
 Mullen, E., 246, 256  
 Müller, W., 348, 560  
 Mulligan, J., 741  
 Murdock, Jr., B.B., 156  
 Mussweiler, T., 627, 678–679  
 Mustard, D.B., 666  
 Mutz, D.C., 258  
 Myers, D.J., 689
- Nabi, R.L., 77  
 Nadler, J., 82, 85, 146n5, 225, 245–246, 258,  
 378–379, 383, 588, 740  
 Nagle, T.T., 274  
 Nalebuff, B., 481  
 Napier, J.L., 765  
 Nario, M.R., 364  
 Nash, J.R., 380–381  
 Navarro-Martinez, D., 303, 318  
 Neale, M.A., 392, 632–633, 690  
 Neate, R., 508  
 Nebergall, R.E., 256  
 Nerenz, D., 256  
 Neter, E., 272, 287n2, 305  
 Newcomb, A., 243  
 Newell, B.R., 10, 175  
 Newport, F., 244  
 Nichols, S., 65
- Nickerson, D.W., 722  
 Nickerson, R.S., 632  
 Nicklisch, A., 49  
 Niedenthal, P.M., 216  
 Niedermeier, K.E., 682  
 Nikiforakis, N., 48, 51  
 Nikitin, M., 138  
 Nisbett, R., 19–20, 447, 769–770  
 Niu, J., 548, 560  
 Nolan, J.M., 257  
 Noll, R.G., 419, 599–600  
 Nordhaus, W.D., 273  
 Norenzayan, A., 51  
 Normann, H., 48, 556, 560  
 Northcraft, G.B., 392, 690  
 Norton, M., 102  
 Nosek, B.A., 684n7  
 Novemsky, N., 393, 394n12, 506  
 Nowak, M.A., 47, 50, 199, 220  
 Nozick, R., 570  
 Nuttin, J., 218  
 Nyborg, K., 37
- Obama, B., 492, 605, 615–617, 734  
 Obama, M., 741  
 Oberholzer-Gee, F., 64, 773  
 O'Brien, B., 113  
 Ockenfels, A., 35  
 O'Connor, M., 530  
 Odean, T., 340, 525  
 O'Donoghue, T., 10, 474, 577n6  
 Oechssler, J., 556  
 Offerman, T., 556  
 Ofshe, R.J., 703  
 O'Gorman, R., 50  
 O'Hara, E.A., 206  
 Okimoto, T.G., 630  
 Oldale, A., 557  
 O'Leary, M., 257  
 Oliar, D., 378  
 Olsen, R.A., 547  
 Olson, M., 45, 257  
 Öncüler, A., 270, 318  
 Önkâl-Atay, D., 337  
 Oosterbeek, H., 32–33  
 Ordeshook, P.C., 37

- Oreg, S., 177  
Orentlicher, D., 99–100  
Orr, D., 627  
Orszag, P.R., 733–734  
Ortmann, A., 135  
Osborne, G., 171  
Oskamp, S., 251  
Osterloh, M., 209  
Ostrom, E., 46  
Oswald, F., 175
- Pace, N.M., 632  
Page, K.M., 199  
Paharia, N., 427  
Paichayontvijit, T., 43  
Papke, L.E., 733  
Parchomovsky, G., 361, 368, 570  
Paredes, T.A., 173, 526–527  
Parfit, D., 288  
Parisi, F., 152, 156n28, 156n32, 284, 398, 504, 606n5  
Park, H.J., 531–532  
Park, R., 711  
Parks, C., 43  
Pasicolan, P., 432  
Paton, B., 740  
Pauer-Studer, H., 85  
Pauly, M.V., 497–499, 502, 511  
Paxton, J.M., 76, 219, 221  
Payne, J.W., 412, 425, 634, 679, 706  
Pearce, J., 259  
Pearse, M., xini  
Peel, E., 369  
Penalver, E.M., 391  
Penner, J.E., 380  
Pennington, N., 667–669  
Penrod, S.D., 672  
Perillo, J., 457  
Perkins, D., 20–21  
Persily, N., 258  
Peruche, B.M., 684  
Pesendorfer, W., 110  
Peters, P.G., 411–412, 419  
Peterson, C.R., 3  
Peterson, D., 258  
Peterson, R., 70
- Pham-Khanh, N., 43  
Phillips, L.W., 254  
Pi, D., 158–159, 164, 398, 504, 606n5  
Piaget, J., 61–63  
Pickel, K.L., 672  
Pierce Wells, C., 256  
Pigou, A.C., 447, 605  
Pildes, R., 102  
Pillutla, M.M., 220n5  
Pirog, R., 766  
Pizarro, D., 76  
Pizer, W.A., 760  
Plant, E.A., 684  
Plato, 119  
Pliske, R., 175  
Plott, C.R., 126, 179, 273, 307–310, 384n3, 388, 601  
Pogarsky, G., 627  
Polinsky, A.M., 327, 395, 432, 576n5  
Politi, D., 243  
Pollock, F., 440  
Polya, G., 10–11, 19  
Popper, K., 601  
Porat, A., 282, 419, 475  
Poser, S., 422  
Posner, E., 31, 45, 207, 225, 412–414, 418, 419n9, 485, 587, 590n17, 752  
Posner, K.L., 356  
Posner, R., 111, 115–117, 125, 168, 182, 196–197, 318, 377, 396, 407–408, 415, 442, 457–459, 491, 521, 523, 553, 569–570, 572, 576, 586, 600, 665  
Potters, J., 556  
Pottow, J., 279, 324, 345  
Poythress, N.G., 366  
Prashker, J.N., 347  
Pratt, J.W., 505  
Prelec, D., 760  
Prentice, R.A., 169–170, 184, 280, 674  
Prescott, E.C., 271  
Preston, C.E., 339  
Preston, E., 364, 636  
Priest, G.L., 284, 505  
Pritchard, A., 527  
Procaccia, Y., 496  
Pronin, E., 147, 152n20, 632–633  
Prosser, M., 655  
Pruckner, G.J., 37

- Puccio, C., 633  
 Puetz, A., 548, 560  
 Pulford, B.D., 345  
 Puto, C., 634  
 Putterman, L., 48
- Qari, S., 604  
 Quattrone, G.A., 429n14  
 Quercia, S., 43-44  
 Quinn, J., 428
- Rabin, M., 18, 35-36, 270, 304, 474, 499,  
 505, 549  
 Rachlinski, J.J., xiin3, 126, 154-155, 164, 167-  
 168, 171-172, 175, 177, 276, 278, 280, 321,  
 326-328, 354, 356, 358-359, 361, 365-366,  
 368, 391-392, 411-412, 419, 465, 544, 561,  
 623, 625, 628, 632-634, 647, 670, 672-673,  
 677-680, 683, 685-686, 690-691, 693,  
 705-707, 710, 714, 760, 762, 766  
 Radcliffe, N.M., 342  
 Radin, M.J., 381-382, 384, 386, 465  
 Raiffa, H., 4  
 Rakos, R., 672, 693  
 Ramsey, F.P., 599  
 Rand, D.G., 41, 47, 50, 220  
 Rapoport, A., 29  
 Rapson, R.L., 755  
 Raseman, S., 252, 758  
 Ratner, R.K., 257, 259, 731  
 Rauscher, F., 570  
 Ravenscraft, D.J., 557  
 Rawling, P., 74  
 Rawls, J., 65, 68  
 Read, D., 12, 18  
 Read, S., 667, 669  
 Reagan, R., 615, 750  
 Rebonato, R., 183n13  
 Reckers, P.M., 365, 690  
 Redelmeier, D.A., 633, 722  
 Redlich, A.D., 133, 137  
 Reed, D., 505  
 Reed III, A., 216  
 Reeve, I., 259  
 Reeves, A.P., 551, 555, 557, 559  
 Regan, D.H., 756
- Reinganum, J., 645  
 Reis-Bergan, M., 342  
 Reiss, S., 176  
 Renner, E., 41, 48-50  
 Revesz, R.L., 751, 759-760, 763  
 Reyna, V.F., 173-174, 675, 678-679  
 Richards, M.J., 665n1  
 Richardson, L.S., 171  
 Richerson, P.J., 29  
 Riedl, A., 33, 48  
 Riis, J., 731  
 Riker, W.H., 37, 97  
 Rinner, W.J., 541, 558-559  
 Risinger, D.M., 714  
 Ritov, I., 66, 68, 71-74, 79, 81-84, 134, 184,  
 271-273, 277-279, 281, 286-287, 316, 326,  
 367, 378, 406, 575, 629, 635, 674, 682,  
 692-693, 707, 709  
 Robben, H.S.J., 605n4  
 Robbennolt, J., 156n30, 278, 341, 425, 627,  
 629-632, 635, 637, 666-667, 680, 703  
 Roberts, A., 686  
 Roberts, M.J., 553  
 Robertson, Q.M., 531-532  
 Robinson, P., 82, 85, 106, 109-110, 183, 245,  
 529, 584  
 Rochberg-Halton, E., 386  
 Rock, E., 208  
 Rockenbach, B., 47  
 Rodriguez, E., 733  
 Roese, N.J., 355, 673  
 Rogers, B.A., 547  
 Rogers, T., 722  
 Röhl, T., 50  
 Romano, R., 522, 524  
 Rose, C.M., 389  
 Rose, M.R., 132  
 Rosenbaum, T., 258  
 Ross, J., 273, 675  
 Ross, L., 152, 152n20, 447, 632-633  
 Ross, L., 252, 769-770  
 Rostechini, A., 587n15  
 Roszkowski, M.J., 547  
 Roth, A.E., 70, 179-181, 391, 649  
 Roth, J., 257  
 Rothbart, M., 581  
 Rothman, A.J., 342



- Rothschild, M., 494–495, 504  
Rottenstreich, Y., 347, 626, 634, 724, 754  
Rowe, R.D., 303  
Rowell, A., 93n1  
Royzman, E.B., 62, 74, 271, 427  
Rozin, P., 251, 271, 722, 725, 735, 764  
Rubin, J.Z., 635  
Rubin, M., 428n13, 429  
Rubin, P.H., 284  
Rubinfeld, D.L., 625, 627  
Rubinstein, A., 289  
Rudman, L.A., 684  
Rudski, J.M., 75  
Ruenzi, S., 548, 560  
Ruhmer, I., 139  
Ruigrok, W., 44  
Russell, G., 765  
Rustagi, D., 44  
Rustichini, A., 204, 378, 545  
Rutherford, M., 328  
Ryan, G., 255
- Sabherwal, S., 525, 548  
Sackett, A.M., 340  
Saez, E., 723  
Sahm, C.R., 739  
Saks, M.J., 623, 671, 676–677, 680–681, 707,  
711, 713  
Sale, H., 525  
Sally, D., 198, 200–202  
Salop, S., 469  
Samuelson, L., 553  
Samuelson, P., 272–273  
Samuelson, W., 300, 393, 674  
Sanchirico, C., 706, 710–711, 713–714  
Sander, F.E.A., 635  
Sanderson, C.A., 257  
Sandroni, A., 504  
Sanitioso, R.B., 339  
Sanna, L.J., 365  
Santora, M., 249  
Santore, R., 276  
Santos, A.C., 179  
Santos, L.R., 314  
Sarat, A., 623  
Sarin, R., 344
- Sausgruber, R., 37  
Savage, L.J., 289  
Sayman, S., 270, 317  
Schacter, D.L., 707  
Schäfer, H.-B., 409–410  
Schauer, F., 692  
Scheffler, S., 286  
Scheier, M.F., 340  
Schelling, T.C., 609  
Schenk, D., 609, 611, 618  
Scherer, F.M., 557  
Schiff, T.M., 556  
Schizer, D., 766n1  
Schkade, D., 82, 102, 155, 168, 412, 424–425,  
575, 591, 600, 676–677, 679, 689–690, 706  
Schlag, P., 687, 769  
Schlanger, M., 107, 422  
Schleifer, M., 72  
Schlingemann, F.P., 557  
Schmid, A.A., 769  
Schmidt, K.M., 35  
Schmitt, E.L., 712–713  
Schmittberger, R., 32  
Schnably, S.J., 382  
Schnall, S., 217  
Schneider, C., 105, 479, 481, 503, 722  
Schönenberger, A., 409  
Schooler, L.J., 181  
Schoorman, F.D., 273  
Schotter, A., 414–416, 418  
Schram, A., 179  
Schrand, C.M., 548  
Schulhofer, S.J., 256  
Schuller, R.A., 672  
Schulze, W., 764  
Schumann, E.L., 681  
Schurr, A., 229, 457n2  
Schut, F.T., 272  
Schwab, S.J., 208, 390, 627  
Schwarcz, D., 493, 496, 503, 509  
Schwartz, A., 278–279, 504, 507  
Schwartz, B., 773–775  
Schwartz, G., 128, 205  
Schwartz, J., 733, 735  
Schwartz, S.H., 37  
Schwarz, N., 365  
Schwarze, B., 32

- Schweder, R.A., 251  
 Schweitzer, M., 272, 674  
 Scotchmer, S., 605n4  
 Scott, R.E., 207, 279, 448  
 Sears, D.O., 257–259  
 Sebenius, J.K., 631  
 Sedikides, C., 339–340  
 Sedlmeier, P., 173  
 Sefton, M., 41, 48–50  
 Segal, D., 501  
 Segal, J., 665n1  
 Segal, U., 345–346, 577, 579  
 Seidmann, D.J., 704  
 Seligman, M.E., 101–102  
 Selten, R., 140, 172  
 Selz, O., 21  
 Semmann, D., 51  
 Sen, A., 600, 761, 771–772  
 Serkin, C., 384–385  
 Sethi-Iyenger, S., 738  
 Shaffer, D.R., 672  
 Shafir, E., 484, 633–634  
 Shalvi, S., 219, 222, 232  
 Shane, F., 577n6  
 Shang, J., 44  
 Shanks, D.R., 175  
 Shanske, D., 599, 602, 609–613  
 Shapira, Z., 559  
 Shapiro, F.R., xin1  
 Shapiro, J., 109  
 Shapiro, M.D., 739  
 Shapiro, S., 387n6  
 Sharkey, C.M., 680  
 Sharot, T., 724  
 Shavell, S., 68, 81, 150n16, 361, 368, 390, 395,  
     407–410, 417–418, 423, 432, 495, 576n5,  
     610–611, 613  
 Shaver, P.R., 219  
 Shaw, B.M., 548  
 Shea, D.F., 272, 483  
 Shefrin, H., 605  
 Shelley, M.K., 760  
 Shepherd, G.B., 632  
 Sheppard, B., 687  
 Sher, S., 182n11  
 Sherif, C.W., 256  
 Sherif, M., 256  
 Sherwin, E., 693  
 Shih, A., 252, 758, 760  
 Shin, J., 630  
 Shinada, M., 48  
 Shleifer, A., 545  
 Shoben, E., 229  
 Shoemaker, P.J.H., 497  
 Shogren, J.F., 305  
 Shor, M., 540, 543, 557  
 Shu, L.L., 217, 227, 356  
 Shultz, T.R., 72  
 Shupp, R., 50  
 Siegel, N.S., 666  
 Siegel, S., 3  
 Siegelman, P., 342, 492, 494, 496, 499–500,  
     505, 509, 512  
 Siegfried, J.J., 553  
 Sigmund, K., 47, 50, 199  
 Silbey, S.S., 623  
 Simmons, C.H., 63  
 Simon, D., 667–669, 693, 709  
 Simon, H., 4, 78, 112n17, 126, 143, 164, 415, 443,  
     555, 645  
 Simon, R.J., 711  
 Simons, D.J., 339–340  
 Simonson, I., 316, 393, 634, 670  
 Sinden, J.A., 273, 302, 306–307, 384, 388  
 Singer, J.W., 380, 396–397, 686  
 Singer, P., 65  
 Singer, R., 722  
 Singh, L., 251  
 Skilton, R., 456  
 Skitka, L., 256–257  
 Skolnick, J.H., 703  
 Slemrod, J., 599, 604, 605n4, 739  
 Sleney, J., 256  
 Slesinger, D., 703  
 Sloan, F.A., 163  
 Slobogin, C., 706  
 Sloman, S.A., 18  
 Sloof, R., 32  
 Slovic, P., 17–19, 79, 167n1, 447, 466n2, 499, 525,  
     724, 753–754, 761, 772  
 Small, D.A., 79  
 Small, D., 772  
 Smith, A., 601  
 Smith, E.R., 396

- Smith, G.E., 274  
 Smith, H.E., 221, 226, 380  
 Smith, H.W., 703  
 Smith, R.J., 175, 684–685  
 Smith, T.W., 244  
 Smith, V., 391  
 Smith, V.K., 163  
 Snelbecker, G.E., 547  
 Snow, C.J., 669  
 Snyder, M., 581  
 Solberg, A.G., 531  
 Songer, D.R., 627  
 Sonnemans, J., 556  
 Sood, S., 634  
 Sorenson, A., 728  
 Sounderpandian, J., 344  
 Sovern, J., 481  
 Spaeth, H.J., 665n1  
 Spalt, O., 271  
 Spector, H., 119  
 Spence, A.M., 505  
 Spence, L.K., 254  
 Spiegelman, E., 37  
 Spiegler, R., 466n1, 467, 471, 475  
 Spier, K.E., 278, 624–625  
 Spitzer, M.L., 20, 155, 306, 328, 393, 421,  
     518–519, 549, 600  
 Spranca, M., 66, 70, 72, 673  
 Sprigman, C., 126, 320–321, 379, 390  
 Squintani, F., 504  
 Srivastava, S., 340  
 Stake, J.E., 285, 314, 389  
 Stalin, J., 79  
 Stallard, M.J., 154n24, 357, 365  
 Stanchi, K., 684  
 Standen, J., 657  
 Stango, V., 722  
 Stanovich, K.E., 15, 171, 215  
 Stark, E., 259  
 Staw, B.M., 273, 675  
 Steblay, N., 138, 672–673  
 Steele, C.M., 768  
 Stein, A., 280, 345–346, 361, 368, 419, 704  
 Stensman, R., 422  
 Stern, S.M., 378–381, 383, 387  
 Sternlight, J.R., 625, 629–630, 632, 634–635, 637  
 Sternquist, B., 274  
 Sternthal, B., 254  
 Steurle, C.E., 607  
 Stigler, G., 31, 542, 555  
 Stiglitz, J., 152, 494–495, 504  
 Stillwell, A.M., 632  
 Stinchcombe, A.L., 244  
 Stocker, S.L., 365  
 Stoebeck, W.B., 395  
 Stone, E.R., 174  
 Stone, R., 480–481  
 Stone II, J., 539–540  
 Stout, L.A., 114, 196–198, 201–202, 204–209,  
     214, 518, 526, 529  
 Strack, F., 678–679  
 Stratmann, T., 132  
 Strecher, V.J., 342  
 Strobel, M., 36  
 Stroop, J.R., 18  
 Stucke, M.E., 539, 551, 555, 557, 559  
 Studebaker, C.A., 425, 627  
 Stultz, R.M., 557  
 Sugarman, D.B., 273  
 Sugden, R., 278, 308–309, 315, 384n3, 413, 500,  
     601n2  
 Sulitzeanu-Kenan, R., 132  
 Suls, J.M., 337–339, 342  
 Sulzle, K., 496  
 Summers, A., 133, 137  
 Sunder, S., 32, 457n2  
 Sunstein, C., xiin3, xin1, 5–6, 15, 19, 23, 28,  
     82, 97–98, 100, 102, 112, 115–117, 125, 150,  
     151n18, 155–157, 162–163, 167–168, 170–172,  
     175–176, 181–184, 213, 227–228, 243,  
     247–248, 250, 283, 289–290, 319, 321, 337,  
     340, 347, 349, 362, 366–367, 379, 390–391,  
     392n9, 398, 412, 418, 424–425, 477–478,  
     479n11, 481, 483–484, 504, 518, 525, 527,  
     530, 533, 539, 571, 575, 583n11, 584, 585n13,  
     590n17, 600, 648, 676–677, 679, 689–690,  
     706, 708–709, 719, 721, 725–726, 729, 731,  
     733, 738, 741, 748, 750, 752–759, 764–765,  
     767, 769  
 Suppes, P., 3  
 Surowiecki, J., 528  
 Surrey, S.S., 290  
 Suslow, V.Y., 555  
 Suter, R.S., 76

- Sutter, M., 50  
 Svenson, O., 147, 152, 474  
 Sydnor, J., 498, 509–511  
 Syverud, K.D., 625, 628  
 Szymanska, E., 67n3, 79, 80n7
- Tabarrok, A., 635  
 Tager, J., 254  
 Talley, E., 306, 345, 390, 518–519, 549  
 Tanford, J.A., 703  
 Tanford, S., 671, 673  
 Tanlu, L., 635  
 Taragin, M.I., 357  
 Taslitz, A.E., 256  
 Tate, G., 526, 548, 557, 560  
 Taylor, D.H., 163  
 Taylor, S.E., 316, 338  
 Teichman, D., 148n9, 154n23, 170, 177, 230,  
     291, 363, 410, 418, 457n2, 460, 527, 580n9,  
     589–590, 605n4, 646, 673, 682, 692–693,  
     707, 709  
 Teitelbaum, J.C., 413–414  
 Tenbrunsel, A.E., 223, 228–229, 231–232  
 Tetlock, P.E., 6, 70, 175, 183–184, 231, 364, 396,  
     684n7, 772  
 Thaler, R., xin1, 5–6, 18, 22–23, 64, 97–98, 100,  
     102, 112, 115–117, 149n13, 167, 170–171, 176,  
     179, 198–199, 213, 224, 227, 268, 270–274,  
     277, 289–290, 300–302, 312, 337, 348,  
     362, 366–367, 384, 388, 391, 392n9, 393,  
     396, 412, 472, 478, 481, 483, 499, 505, 518,  
     525, 539, 549, 554, 571, 584, 590–591, 593,  
     599–600, 601n1, 605–608, 612, 647–649,  
     674, 706, 709, 719, 726, 731, 752, 754,  
     756–757, 769  
 Thayer, J.B., 439  
 Theiss-Morse, E., 258n5  
 Thevenot, D., 765  
 Thibaut, J., 396  
 Thomas, R.S., 208  
 Thompson, L., 224, 270, 338, 626, 632  
 Thompson, W.C., 681  
 Thöni, C., 41–45, 49, 51–53, 57  
 Thuronyi, V., 290  
 Tian, K., 386n5  
 Tillers, P., 711
- Tindale, R.S., 550  
 Titmus, R., 587  
 Todd, P.M., 12, 761  
 Togler, B., 604  
 Tomlinson, T.M., 672  
 Tong, K.K., 630  
 Tontrup, S., 184, 316–317, 323  
 Topa, G., 723  
 Toplak, M.E., 15  
 Tor, A., 125, 345, 362, 378, 524, 526, 530, 539,  
     541, 543–544, 549, 553–554, 556, 558–559,  
     561, 649  
 Torgler, B., 44  
 Traulsen, A., 50  
 Trautmann, S.T., 344–345  
 Traxler, C., 44  
 Trebilcock, M., 442–443  
 Treitel, G.H., 369  
 Treviño, L.K., 214, 232  
 Tribe, L.H., 709, 771  
 Trivers, R., 218  
 Trope, Y., 348  
 Trossen, D.R., 275, 465, 472  
 Trubeck, D.M., 624  
 Turiel, E., 62–63  
 Turner, M.A., 368  
 Tur-Sinai, O., 390, 398  
 Tversky, A., 3–4, 9, 11, 19, 28, 79, 96, 112, 126, 143,  
     148, 149nn12–13, 154, 163, 167, 169, 171–174,  
     178–181, 213, 268, 270–272, 277, 305–309,  
     312, 345, 347–348, 413, 445–447, 466n2,  
     500, 506, 525, 545, 547, 581, 584, 599, 601n1,  
     609, 626–628, 634, 647, 670, 674, 678,  
     681–682, 703–705, 723–724, 753, 760  
 Tyler, T.R., 224, 246–247, 396, 630  
 Tyran, J.-R., 42–43
- Ubel, P.A., 80  
 Uhlich, G.R., 140  
 Uhlmann, E.L., 217  
 Ulen, T.S., 97, 108, 110, 115, 167, 213, 270, 366,  
     391, 395, 410, 412, 461, 518, 527, 568n1, 624,  
     649, 709, 711, 713  
 Ungar, L., 76  
 Unger, P., 773  
 Updegraff, J.A., 179n9

- Van Aaken, A., 773  
Vanberg, C., 457n2  
Van Damme, E., 560  
Van de Kuilen, G., 32  
Vandenbergh, M., 203–204, 758  
Van den Bos, K., 219  
Van der Walt, N., 531  
Van Dijk, E., 304, 316, 388, 393  
Van Geen, A., 227  
Van Harreveld, F., 630  
Van Kesteren, J., 690  
Van Knippenberg, D., 304, 316, 388, 393  
Van Koppen, P.J., 628  
Van Lange, P.A.M., 29, 32, 35, 43, 48  
van Praag, M., 532  
Van Vugt, M., 50  
Vars, F.E., 152n21, 287, 291, 367, 406, 666, 681, 709  
Vatn, A., 770  
Vazquez, J.M., 604  
Veblen, T., 767  
Velleman, J.D., 85  
Vermeule, A., 585n13  
Vertova, P., 37  
Vesterlund, L., 306  
Viard, A.D., 276  
Vick, D., 660  
Vidmar, N., 632, 666, 677, 692–693  
Vieider, F.M., 344–345  
Viki, G.T., 256  
Villegas-Palacio, C., 43  
Vinnicombe, S., 531  
Viscusi, W.K., 205, 345, 360, 412n4, 419, 445–446, 673, 679, 690n9, 691, 751–752, 760, 762  
Visser, P.S., 255  
Vlek, C., 760  
Voci, A., 429  
Vohs, K.D., 216, 355, 673  
Volk, S., 44  
Von Hippel, W., 218  
  
Wade, M., 644  
Wagenaar, W.A., 626  
Wagner, J., 414  
Wakker, P., 344–345, 348  
Wakslak, C.J., 247  
Walker, J.M., 46, 50  
Walker, L., 396, 733  
Walster, E., 63, 257  
Walster, G.W., 63  
Walters, S., 733, 735  
Wambach, A., 496  
Wansink, B., 735  
Ward, A., 632  
Ward, P., 6  
Warren, E., 341  
Wasserman, J., 248–249  
Waters, N., 624  
Watson, T.D., 102  
Weaver, R., 314, 316  
Weber, M., 344–345, 348, 548, 556  
Weber, R.A., 222, 231  
Weinrib, E., 586  
Weinshall-Margel, K., 665  
Weinstein, N.D., 147n7, 162, 337–339, 342, 378, 396, 445, 474, 646, 724  
Weintraub, J., 456  
Weisbach, D.A., 395  
Weisbach, M.S., 550  
Weisberg, M., 117n21  
Weisel, O., 49  
Weiss, W., 254  
Wellborn III, O.G., 703  
Wellman, H.M., 22  
Wells, G., 138, 679, 681–682  
Wells, M.T., 666, 677, 679, 706  
Wendinga, R.M., 760  
Wenglert, L., 147n7  
Wengström, E., 42–43  
Werden, G.J., 540–541, 543, 557  
Werdigier, J., 531  
Werner, C.M., 250  
Wertheimer, M., 16  
West, R.F., 15, 171, 215, 242  
Whitman, D.A., 395  
Wiener, J.B., 751  
Wilcox, R.T., 22  
Wilcoxon, P.J., 770  
Wilde, L.L., 504, 507  
Wilkins, P.A., 686  
Wilkinson, N., 35  
Wilkinson-Ryan, T., 182n11, 184, 457n2, 460–461  
Willham, C.F., 355, 359  
Williams, D.R., 328

- Williams, S.H., 147, 337, 340, 343, 362, 446,  
     474, 499, 585, 646, 692, 724  
 Williams, V., 454  
 Williamson, O.E., 443, 546  
 Willis, L., 103, 106, 326, 503  
 Wilson, M.A., 771  
 Wilson, T.D., 678  
 Windschitl, P.D., 338, 340  
 Winkler, A., 244  
 Winter, R., 520  
 Wissler, R., 671, 676, 679–681  
 Wistrich, A.J., 175, 278, 280, 358–359, 628,  
     632–634, 670, 672–674, 678, 680, 683,  
     690–691, 693, 706–707, 710, 714  
 Witte, A.D., 257  
 Wittermoth, S., 633  
 Wittman, D., 416–417  
 Wladis, J.D., 453–454  
 Wlodarski, R., 339  
 Wong, W., 310, 388n7, 601n2  
 Woo, C., 303  
 Woocher, F.D., 703  
 Wood, G.C., 429n14  
 Woods, Susan, 256  
 Worthington, D.L., 154n24, 357, 365, 708  
 Wright, E.W., 687–688, 692  
 Wright, J.D., 539–540  
 Wright, R., 653  
 Wrightsman, L.S., 156  
 Wuthnow, R., 259  
  
 Yamagishi, T., 44, 46, 48, 50  
 Yan, J., 526  
  
 Yang, Q., 216  
 Yang, Y.T., 128  
 Yaniv, G., 580  
 Yoeli, E., 51  
 Yosifon, D., 533, 768  
 Young, D., 685  
 Young, L., 73, 219  
 Young, M.J., 633  
  
 Zak, P.J., 210  
 Zakaras, L., 632  
 Zamir, E., 23, 42, 71, 97, 101, 103, 115n19, 117n21,  
     134, 148n10, 154n23, 184, 271, 277, 281,  
     283–287, 291, 321, 323–324, 326, 367, 384,  
     406, 421, 446–447, 484, 506, 556, 590n16,  
     608, 618, 629, 635, 647–648, 658, 666, 674,  
     682, 692–693, 707, 709, 721, 723, 726, 733,  
     736, 754, 766, 771, 1125  
 Zarsky, T.Z., 275, 486  
 Zauberman, G., 481  
 Zechman, S.L., 548  
 Zeckhauser, R., 272, 300, 393n10, 496–498, 511,  
     601, 608, 674  
 Zeiler, K., 126, 273, 307–310, 312, 384n3, 388, 601  
 Zeisel, H., 651  
 Zelinsky, E.A., 290  
 Zelizer, V.A., 396  
 Zhang, B., 50  
 Zhong, C.B., 217  
 Zhurakhovska, L., 560  
 Zimbardo, P., 220n5, 588  
 Zinman, J., 722  
 Zwane, A.P., 723





## SUBJECT INDEX

---

- abortion, 73–74, 78, 246
- active choice requirements, 737–738. *See also*
  - nudges
- actively open-minded thinking, 21
- adolescent risk taking, 181
- affect heuristic, 17, 335, 340
- affective forecasting
  - basic properties of, 101–102
  - criminal law and, 108–110
  - endowment effect and, 322
  - set point of individual happiness and, 101–102
  - tort law and, 107–108
- Affordable Care Act
  - breastfeeding and, 242
  - disclosure requirements in, 728–730
  - employer mandate and, 734
  - insurance law and, 491–492, 495, 513
  - National Federation of Independent Businesses v. Sebelius* and, 254
- Allais paradox, 346
- altruism, 32, 79–80, 117, 200
- ambiguity aversion
  - contract law and, 345, 349
  - definition of, 336
  - gains *versus* losses and, 344–345
  - plea bargaining and, 345–346
  - risk aversion and, 345
  - subjects' perceived competence and, 345
  - taxation and, 346
- anchoring effects
  - judicial decision-making and, 678–680
  - litigation and settlement and, 627, 636
  - plea bargaining and, 657
- antitrust law
  - aftermarket power and, 551–552
  - behavioral law and economics research and, 550–561
  - bundling and, 552
  - cartels and, 555–556
  - firms' behavior and, 545–550
  - firms' learning and, 545–546
  - market behavior and, 542–546
  - market entry and, 552–554
  - mergers and, 541, 548–549, 553–554, 556–557, 559–560
  - overoptimism bias and, 553, 557
  - predation and, 541, 544, 554
  - rational choice theory and, 540–541
  - resale price maintenance (RPM) and, 557–559
  - selection mechanisms and, 544–545
  - tying and, 541, 543–544, 551–553
  - unexploited market power and, 554–555
- Aronson v. Lewis*, 529
- Asian Disease problem (Kahneman and Tversky), 79, 154–155
- AT&T Mobility v. Concepcion*, 478
- attribute substitution heuristic, 17
- attribution bias
  - evidence law and, 711, 713
  - overoptimism bias and, 337–339, 524, 711, 713, 769–770
- availability heuristic
  - contract law and, 446, 449–450, 452
  - criminal law and, 583–585
  - environmental law and, 762
  - evidence law and, 704–706, 713
  - insurance law and, 501
  - litigation and settlement and, 627
  - tort law and, 414, 418–419
- Bankruptcy Abuse Prevention and Consumer Protection Act (BAPCPA), 105
- Bayesian probability theory, 3–4, 667, 669
- behavioral ethics
  - automaticity of morality and, 219

- behavioral ethics (*cont.*)
- automaticity of self-interest and, 215–218
  - automatic reasoning and, 220–221, 227, 231
  - awareness and, 218
  - blind spots and, 231–232
  - bounded ethicality and, 222–224
  - conflicts of interest and, 230–231
  - debiasing and, 228–229
  - disambiguation and, 230–231
  - egotism and, 218
  - embodiment and, 216–217
  - ethical fading and, 223
  - fairness and, 224, 226
  - limitations of, 213, 226
  - limited-awareness problem and, 231
  - micromanagement and, 233
  - moral cleansing and, 224
  - moral disengagement and, 223–224
  - moral hypocrisy and, 221, 223
  - moral licensing and, 223–224
  - motivated reasoning and, 217, 220, 222
  - normative implications of, 226–229
  - nudge approach and, 227–228
  - physiological research and, 217, 221n6
  - priming and, 216–218, 228
  - prosocial behavior research and, 214
  - reflection and accountability and, 231–232
  - regulating situations and, 231–232
  - self-concept maintenance and, 222–223
  - self-interest and, 214–226
  - shaming and, 225
  - social norms and, 224–225
  - statistical unethicality and, 229–230
  - two-systems theory and, 215–216, 219–220, 227, 230–231
- Behavioral Insights Team (United Kingdom), 481, 483
- behavioral law and economics (BLE). *See also* empirical methods in behavioral law and economics
- analytical methods in, 177–182
  - biased-samples perspective and, 173
  - commitment to empirical evidence and, 171
  - consequentialist approach and, 183–184
  - domain-specific phenomena and, 172, 177, 184
  - dread risks and, 170
  - economists' criticisms of, 112–115
  - fuzzy-trace theory and, 173–174
  - goal-based decision-making and, 175–176
  - law and economics research and, 167–172, 175
  - legal academic criticisms of, 115–119
  - motivational assumptions in, 175–177
  - naturalistic decision-making and, 175
  - normality bias and, 184
  - personal autonomy–based criticisms of regulations based on, 119–120
  - prescriptive interventions and, 170–172, 179, 181–183
  - rational choice theory challenged by, 93, 95–97, 101–102, 111–114, 116, 143, 164, 167–172, 175, xi
  - rationality assumptions in, 168–175
  - rejections of efficiency in, 183
  - tautological problems with, 117, 197
  - utility maximization and, 182–184
- Belgium, 100
- benevolent biasing, 158–161, 164
- bias, definition of, 5, 145. *See also specific biases*
- biasing. *See also* debiasing
- benevolent biasing and, 158–161, 164
  - cognitive leveraging and, 158, 161–164
  - deliberation costs and, 159–160
  - double biasing, 162
  - environmental law and, 160–162
  - misinformation campaigns and, 159–160
  - second-order effects and, 160
- BillShrink.com, 482
- Bite Club* (video game), 106
- blood donation, 250
- boards of directors
- diversity and, 530–533
  - groupthink and, 530, 532
  - independence of, 529–531, 533
  - in-group *versus* out-group dynamics in, 529, 533
  - Sarbanes-Oxley Act and, 151n19, 157, 533
  - women and, 530–531
- bounded willpower, 117, 146, 648–649, 660
- Brady v. Maryland*, 656
- breastfeeding, 242
- broken windows theory, 588
- Brooke Group Ltd. v. Brown & Williamson Tobacco Corp.*, 540–541

- bundling, 552  
 burden of proof  
   endowment effect and, 326–327  
   hindsight bias and, 367  
   loss aversion, 280–281  
   varieties of, 134, 280–281  
 Business Judgment Rule, 368
- calibration, 8  
 capital punishment, 255  
 cartels, 555–556  
 certainty effect  
   Allais paradox and, 346  
   circumstantial evidence and, 336, 348  
   Clean Air Act and, 349  
   common ratio effect and, 347  
   definition of, 344  
   delayed payouts' impact on, 347–348  
   emotionally salient outcomes and, 347, 349  
   environmental law and, 348–349  
   probabilistic insurance and, 348  
   Superfund site cleanups and, 348  
 Charlotte (North Carolina), 254  
*Chicago Tribune*, 243  
 Children's Health Insurance Program, 734  
 Choquet Expected Utility Theory, 413  
 cigarette smoking  
   change in social meaning of, 85, 248–249, 251, 740  
   moral attitudes and, 242, 248–252  
   overoptimism bias and, 162–163, 341–342, 724  
   second-hand smoke and, 248–249, 251, 255  
 Clayton Act, 541  
 Clean Air Act, 349  
 Coase Theorem, 97, 301–302, 310–311, 313, 329  
 cognitive leveraging, 158, 161–162, 164  
 Community Living Assistance Services and Supports Act (CLASS Act), 734  
 Competition Commission (United Kingdom), 501, 505, 508  
 conformism, 44  
 congruence heuristic, 14  
 consequentialism, 67, 182–184, 288. *See also* utilitarianism  
 Consumer Financial Protection Bureau (CFPB), 107, 479, 481, 484, 731
- consumer transactions. *See also* contract law  
   add-on products and, 470  
   annual cost-of-ownership information and, 481  
   behavioral economics theory of, 466–474  
   behavioral market failure and, 467–468  
   collective-action problems and, 469–470, 496  
   competition's role in, 468–469  
   comprehensive disclosures for sellers and, 482–483  
   consumer misperceptions and, 466, 468–472, 475–477, 480, 487  
   contract complexity and, 471–473  
   default rules and, 483–484  
   deferred costs and, 473–474  
   disclosure of product-attribute information, 479  
   disclosure of product use regulation and, 480–481  
   distorted competition and, 475–476  
   distributional concerns and, 476–477  
   hindered competition and, 475  
   imperfect information and, 467–468, 479  
   imperfect rationality and, 467–476, 479, 481–482, 486  
   learning from mistakes and, 470–471  
   mandatory rules and, 477–478  
   market corrections and, 469–471  
   myopia and, 474, 482  
   overoptimism bias and, 468, 474, 482  
   perceived net benefits *versus* actual net benefits and, 467–469  
   right to withdraw provisions and, 484–486  
   safe harbors and, 484  
   simple disclosure for consumers and, 481–482  
   teaser-rate contracts and, 468  
   total-cost-of-ownership (TCO) information and, 481  
   welfare implications of, 475–477  
 contract law. *See also* Consumer transactions  
   acceptance by mail and, 438–439  
   availability heuristic and, 446, 449–450, 452  
   behavioral law and economics discipline and, 442–447

- contract law (*cont.*)
- bounded rationality and, 442–443
  - “Chicago” law and economics discipline and, 441–442
  - cognitive psychology and, 445–447
  - defects in capability and, 446–447
  - defects in cognition and, 450–453
  - defects in disposition, 445–446
  - deferred exchange and, 206
  - expectation damages and, 454, 458–461
  - express conditions and, 450–453
  - forfeiture principle and, 450–451, 462
  - formalism and, 438–441
  - form contracts and, 455–456
  - informed parties assumption and, 442–443
  - in-kind relief *versus* monetary relief, 459, 461
  - liquidated damages and, 448–450
  - loss aversion and, 275, 446–447, 454–455
  - modern contract law and, 441
  - moral considerations and, 459–460
  - mutual mistakes and, 453
  - option theory of contracts and, 457–459
  - overoptimism bias and, 445–446, 449, 452, 455
  - punitive damages and, 460–461
  - rational choice theory and, 439, 441–446
  - rational ignorance and, 443, 455
  - relational contracting and, 206–207
  - resale paradigm and, 458–459
  - social propositions’ role in, 438–441
  - tacit assumptions and, 444–445
  - telescoping faculties and, 447, 449
  - theory of efficient breach and, 457–459
  - unconscionability and, 448, 451
  - unexpected circumstances and, 453–455
- corporate and securities law
- academic defenses of, 521–523
  - assumption of market accuracy in, 526
  - boards of directors and, 151, 528–533
  - confirmation bias and, 526
  - corporate governance and, 519, 523–528, 530
  - corporate managers and, 519–522, 524–528
  - disclosure requirements and, 518, 525–527, 534
  - efficient market assumptions and, 521
  - fiduciary duties and, 208, 518–520, 522–523, 527, 529, 533–534
  - hindsight bias and, 357, 527
  - judicial deference and, 520–521
  - overoptimism bias and, 525
  - prosocial behavior and, 207–209
  - rational choice theory and, 518–521, 523–524, 526–527, 534
  - Sarbanes-Oxley Act and, 151n19, 157, 533
  - sell-side securities analysts and, 525
  - shareholders and, 519–522, 526–527, 529, 533
  - voluntariness and, 519–523
- credit cards. *See also* consumer transactions
- complexity of contracts for, 471
  - Credit Card Accountability Responsibility and Disclosure Act of 2009 (CARD Act), 477, 483–484
  - credit limits and, 483–484
  - deferred costs and, 473–474
  - disclosure requirements and, 480–481
  - overoptimism bias and, 341, 470
  - perceived net benefit *versus* actual net benefit and, 467–468
  - regulation of late fees for, 477
- criminal law
- actual punishment and, 572, 583
  - affective forecasting and, 108–110
  - availability heuristic and, 583–585
  - behavioral economics research and, 574–590
  - bounded rationality and, 571
  - broken windows theory and, 588
  - communicative theory and, 571
  - criminals’ risk propensities and, 573–574, 576, 579–581, 583–585
  - crowding out behavior and, 587
  - decision utility, 590–592, 594
  - deterrence and, 108–110, 158–160, 570, 572, 575–577, 579–581, 585, 591–592, 594
  - discounting of the future bias and, 573, 577, 579
  - distributive justice and, 569–570
  - duty of rescue and, 586–587
  - expected punishment and, 572–573, 575–579, 585, 588
  - experienced utility, 590–592, 594
  - happiness research and, 590–594
  - hindsight bias and, 358–359
  - law and economics research and, 570–576

- neoclassical economic analysis of, 568–570  
 overoptimism bias and, 585–586  
 perceptions regarding justness of law and, 588–589  
 positive duties and, 586–587  
 prediction *versus* postdiction in, 581–583  
 probability of detection considerations and, 109–110, 158–160, 568, 571–573, 575–579, 583–588, 590  
 prospect theory and, 579–581  
 public interest and, 594–595  
 rational choice theory and, 571–574, 589  
 retributive justice and, 569–570, 590, 592–594  
 rules *versus* standards in, 582–583  
 sanction size and, 108–110, 158, 568, 572–573, 575–578, 583–585, 587–588, 592–593  
 social norms theory and, 587–588  
 subjective disutility and, 591–594  
 uncertainty effect and, 577–580
- date rape, 256
- Daubert v. Merrell Dow Pharmaceuticals*, 708
- Debiasing. *See also* biasing  
 adjudicative forms of, 153–154, 156, 163  
 agency law and, 157  
 behavioral ethics and, 228–229  
 blind spot bias and, 152  
 changing preferences and, 145–147  
 corporate law firms and, 156  
 deliberation costs and, 153, 159–160, 162  
 endowment effect and, 157  
 framing effects and, 154–155  
 incentive alignment and, 147, 153, 162  
 insulation and, 149–152, 157, 164  
 juries and, 154–156  
 litigants and, 156  
 market failures and, 158, 161  
 search costs and, 153  
 self-correction and, 152–153  
 substantive forms of, 153–154, 156  
 thin *versus* thick rationality and, 146–147
- default rules  
 active choice requirements as alternative to, 737–738  
 automatic enrollments and, 721–722, 733–737  
 basic properties of, 97–98  
 consumer transactions and, 483–484  
 endowment effect and, 304, 323–325  
 environmental law and, 748, 757–758, 769  
 health care and, 728–729, 734  
 loss aversion and, 279–280  
 as means of combating inertia, 735–736  
 nudges and, 721, 733–739  
 organ transplants and, 98–100, 272  
 retirement savings and, 733–734  
 school meals and, 734
- Delaware Chancery, 208
- deontology, 75–78, 286–288
- Department of Agriculture (USDA), 728
- Department of Education, 732
- Department of Labor, 732–733
- descriptive theory of decision-making, 4–5, 9–10
- deterrence  
 criminal law and, 108–110, 158–160, 570, 572, 575–577, 579–581, 585, 591–592, 594  
 distribution of goods and bads and, 82  
 tort law and, 81, 409, 413–415, 418–420, 423–424, 432
- diabetes, 132
- dictator game  
 basic rules in, 32  
 equity theory and, 64  
 heterogeneity of outcomes in, 177–178  
 prosocial behavior and, 198–200  
 similarity to no-liability regimes and, 415n7
- disclosure requirements  
 clarity and, 729  
 competition and, 731–732  
 summary disclosure *versus* full disclosure, 727, 732–733
- discounted utility, 7–9
- Discover Bank v. Superior Court*, 478
- distribution of goods and bads, 80–82
- docility, 78
- Dodd-Frank Financial Reform Act, 477, 481, 484, 731
- “do no harm” heuristic, 14, 17
- Doorways to Dreams (D2D) Fund, 106
- double biasing, 162

dread risks, 170

drunk driving, 583

*Eastman Kodak Co. v. Image Technical Services*, 551

Emergency Planning and Community

Right-to-Know Act of 1986, 732

eminent domain, 378, 381

empirical methods in behavioral law and economics

causality and, 132–133

design economics and, 180

difference estimation and, 132–133

endogeneity problems and, 133

external validity and, 129, 131, 135, 138, 179–180

field evidence and, 130–133

instrumentation and, 133

internal validity and, 129, 135–136, 138

lab research and, 130–131, 135–136

meta-analysis and, 138

qualitative methods' advantages and, 137

quasi experiments and, 140

replication of findings and, 133–134, 138, 179, 181

simulation and, 137

survey research and, 130–131, 133–134

vignette studies and, 130–131, 134

endowment effect

affective forecasting problems and, 322

agents *versus* principals and, 306–307, 323

alternative endowments and, 304–305

among nonhuman primates, 314

assigning entitlements and, 318–321

attachment to substantive endowments and, 313–314

burden of proof and, 326–327

causes of, 311–317

Coase Theorem and, 300–302, 311, 313, 329

contingent valuation surveys and, 301–302

controlled experiments demonstrating, 302–303

creators *versus* noncreators and, 320–321

criticism of, 113–114

debiasing and, 157

default rules and, 304, 323–325

economic analysis of law and, 318–328

environmental entitlements and, 315, 327–328

environmental law and, 754, 764, 766–767

evolutionary theory and, 313–314

facilitating private interactions and, 323–326

goods held for exchange and, 114

laboratory experiment practices and, 307–310

legitimacy norms and, 315

liquidity constraints and, 312

loss aversion and, 149, 268, 270, 272–273, 275, 291, 312–317

market mimicking entitlement

assignments and, 318–319

market participation's impact on, 305–307

market substitutes and, 305

nonrivalrous endowments and, 320–321

nudges and, 723

offer-asking gap and, 301

plea bargaining and, 648, 659

pricing experiments and, 307–309

property law and, 384–385

property rules *versus* liability rules and, 327

query theory and, 317

redistributing rights and, 321–323

reference points and, 303–304

regret avoidance and, 315–317

regulatory inertia and, 321

rules protecting entitlements and, 326–328

taxation and, 600–601, 618

trading experiments and, 309–310

transactional disutility, 314–315

wealth effects and, 311–312

willingness to accept (WTA) and, 270,

273, 302–303, 305–309, 311–316, 318–322, 324–328

willingness to pay (WTP) and, 270, 273,

302–303, 305–309, 311–316, 318–322, 324, 326, 328, 769

environmental law

affect bias and, 753–754

availability cascades and, 753, 760

availability heuristic and, 762

biasing and, 160–162

- “cancer premium” and, 762–763  
 certainty effect and, 348–349  
 coal ash regulations and, 763–764  
 cognitive dissonance and, 768  
 commensuration and, 751  
 cost-benefit analysis (CBA) and, 748,  
     750–753, 759–760, 763, 769–770, 773–775  
 default rules and, 748, 757–758, 769  
 disclosure requirements and, 748, 757–758  
 dissonance reduction and, 765  
 early rationalist critiques and, 750–752  
 ecosystem complexity and, 770  
 endogeneity and, 749, 767, 774  
 endowment effect and, 754, 764, 766–767  
 energy efficiency and, 160–162, 758, 762  
 foreground bias and, 771  
 framing and, 760, 769  
 gasoline taxes and, 766  
 greenhouse gas regulation and, 321, 733,  
     757, 766  
 group polarization and, 753, 755, 757  
 individual *versus* statistical victims and,  
     772  
 information salience and, 753–754, 757  
 libertarian paternalism and, 756–758,  
     769–771  
 loss aversion and, 754, 764, 766–767  
 market pricing and, 772–774  
 monetary value of a statistical life (VSL)  
     and, 752  
 moral panics and, 753  
 natural disasters’ impact and, 759–761, 766  
 nudges and, 749, 756–758  
 precautionary regulation and, 749–750,  
     753–755, 759–761, 767  
 probability neglect and, 753  
 prospect theory and, 760  
 rational choice theory and, 759, 767  
 regulatory impact analysis and, 750–751  
 risk perception and, 748–749, 753–755,  
     759–760, 762  
 risk regulation and, 754, 762–763, 774  
 shoves and, 749, 764–767  
 social norms and, 757–758, 775  
 status quo bias and, 753–754, 757, 765–767  
 stigma effect and, 763–764  
 Superfund sites and, 764  
 willingness to pay (WTP) evaluations and,  
     769  
 Environmental Protection Agency (EPA)  
     “cancer premium” calculation and,  
         762–763  
     Clean Air Act and, 349  
     coal ash regulation and, 763–764  
     fuel efficiency disclosure requirements  
         and, 730, 732, 757  
     greenhouse gas reporting rule and, 733  
 equality heuristics, 80–81  
 equal weighting heuristic, 13  
 equilibrium analysis, 504, 769  
 European Commission, 720, 751  
 European Union, 272, 544, 619, 720  
 euthanasia, 71, 102  
 evidence law  
     adversarial process’ impact on bias and,  
         705–706  
     attribution errors and, 711, 713  
     availability heuristic and, 704–706, 713  
     character evidence and, 711–714  
     circumstantial evidence and, 707  
     expert testimony and, 707–708  
     eyewitness testimony and, 707  
     hearsay and, 704–705  
     hindsight bias and, 706, 708, 710–711  
     omission bias and, 709  
     overcorrection bias and, 704  
     proposals to change, 706  
     rational choice theory and, 703, 709  
     representativeness heuristic and,  
         705–707, 713  
     sex crimes and, 705, 711–713  
     simulation heuristic and, 707  
     standards of proof and, 708–709  
     statistical data and, 708  
     subsequent remedial measures and,  
         710–711  
 evolutionary biology, 117  
 evolutionary psychology, 9, 287  
*Ewing v. California*, 653  
 Executive Order 12291 (Reagan), 750  
 expectation damages, 454, 458–461  
 expected utility theory  
     defining elements of, 7  
     deviations from, 9



- expected utility theory (*cont.*)  
  insurance law and, 492–494, 498–500,  
    511, 513  
  loss aversion, 268–270, 273, 288–289  
  probability and statistics in, 8  
extended warranties, 505–509
- Family and Medical Leave Act, 252  
fast and frugal heuristics, 12, 172–173  
Federal Cigarette Labeling and Advertising  
  Act, 248  
Federal Communications Commission  
  (FCC), 481, 483  
Federal Public Health Cigarette Smoking Act  
  of 1969, 248  
Federal Reserve Board, 479  
*Federal Rule of Evidence* 403, 710  
*Federal Rule of Evidence* 413, 711  
Federal Trade Commission (FTC), 478–479,  
  485, 736  
financial literacy education, 103–107  
firms  
  corporate boards and governance in,  
    549–550  
  institutions within, 548–550  
  learning by, 545–546  
  managers in, 547–548  
  organizational repairs in, 548–549  
  overoptimism bias and, 525, 548, 553  
  rational choice theory and, 546–550  
first-order thoughts, 118  
fixed pie bias, 632  
Flood Control Act of 1936, 750  
Food and Drug Administration (FDA), 731  
Food Marketing Institute, 741  
framing effects  
  basic properties of, 96–97, 148–149  
  debiasing and, 154–155  
  environmental law and, 760, 769  
  litigation and settlement and, 628  
  nudges and, 722  
  plea bargaining and, 658  
  taxation and, 599–600, 609, 614  
  tort law and, 420–421  
fuel efficiency, 758  
fuzzy-trace theory, 173–174
- Geneva Papers on Risk and Insurance*  
  (journal), 491  
*Gerhardt v. Continental Insurance Cos.*, 456  
gift-exchange game, 33–34  
*The Gift Relationship: From Human Blood to*  
  *Social Policy* (Titmus), 587  
gist memories, 173–174  
*Goldberg v. Kelly*, 283  
Golden Rule, 421  
Good Samaritan laws, 245, 258, 586  
*Graham v. John Deere Co.*, 358  
Green Behavior, 720  
greenhouse gas regulation, 321, 733, 757, 766  
*Griffith v. Brymer*, 453  
*Griggs v. Duke Power Co.*, 229  
Grocery Manufacturers Association, 741  
gun control, 243–245, 252–253
- Hadley v. Baxendale*, 370, 458  
*Haifa Municipality* (case), 431  
happiness research, 101–102, 590–594  
harm principle, 255  
heuristics. *See also specific heuristics*  
  Bayesian probability theory and, 3–4  
  children’s learning and, 15–16  
  debiasing and, 19–20  
  early history of research on, 10–14  
  isolation effects and, 17–18, 83  
  overgeneralization within, 16–17  
  representativeness and, 3, 11  
  two-system theory and, 18–19  
  undergeneralization within, 16  
hindsight bias  
  basic properties of, 148  
  bifurcation of legal proceedings as a means  
    of dealing with, 366–367  
  British-Gurka war experiment and, 355  
  burden of proof and, 367  
  consider-the-opposite strategy and,  
    364–365  
  contracting out of, 369–370  
  corporate law and, 357, 527  
  countervailing factors and, 362–363  
  criminal law and, 358–359  
  debiasing and, 364–366  
  dividend payments and, 369

- experiments demonstrating, 355, 358, 367  
 impact on production of evidence, 356–360  
 judges and, 359–360, 370  
 in judicial decision-making, 356–360, 673  
 jurors and, 356, 360, 365–367, 370  
 motivated reasoning and, 363  
 normative implications of, 354, 360–363  
 overoptimism bias and, 362–363  
 patent law and, 357–358  
 substantive legal rule changes as means of  
 dealing with, 367–369  
 tort law and, 356–357, 410–412, 414, 419  
*Holiday Inns of America Inc. v. Knight*, 451  
*homo economicus*. *See also* rational choice  
 theory  
 criticism of the model of, 30, 33–34, 43,  
 46, 53, 197, 199–200, 202–204, 208–210,  
 752–753  
 Global Financial Crisis (2008–2009) and,  
 518  
 history of the idea of, 31, 195–196  
 as sociopathic actor, 196–197  
 utility maximization and, 31–33, 195–197, 206  
*homo reciprocans*, 30–33, 35, 46  
 human rights, 282–284  
 Hurricane Katrina (2005), 765  
 Hurricane Sandy (2012), 766  
 hyperbolic discounting, 10, 104, 116, 577
- indirectness bias, 73–74  
*In re Cellphone Termination Fee Cases*, 477  
 insurance law  
 adverse selection problem and, 492,  
 494–496, 502, 504, 512  
 Affordable Care Act and, 491–492,  
 495, 513  
 availability heuristic and, 501  
 behavioral economics and, 497–504  
 complexity aversion and, 501  
 consumer heterogeneity and, 471, 478,  
 503–504  
 deductibles and, 494–495, 498, 504–505,  
 509–512  
 demand-side anomalies and, 498–502  
 disclosure and, 496, 503, 508–509  
 economic basis of, 493–494, 497  
 employer-provided health insurance and, 512  
 excessive discounting and, 499  
 expected utility theory and, 492–494,  
 498–500, 511–513  
 extended warranties and, 505–509  
 government regulation and, 493–497, 502,  
 508–509, 511–513  
 hyperbolic discounting and, 501  
 imperfect rationality and, 502–504  
 information problems facing consumers,  
 496–497  
 information problems facing insurers and,  
 494–496  
 information processing problems  
 and, 501  
 loss aversion and, 500, 502–503, 505–506,  
 509–510  
 low-deductible home owners' insurance  
 and, 509–511  
 market equilibria and, 504–505, 507–513  
 moral hazard and, 494–496, 498, 502  
 overoptimism bias and, 499, 504  
 regret aversion and, 500, 502–503, 505–506,  
 509, 512  
 risk aversion and, 494–495, 498–499, 505,  
 509–511  
 risk classification and, 495  
 threshold effects and, 500–501  
 tort law and, 410, 417–418, 425  
 underwriting and, 495  
 International Association for the Study of  
 Insurance Economics, 491  
 International Monetary Fund, 619  
 Iowa Civil Rights Act, 242–243
- Japan, 43  
 Jekyll-Hyde syndrome, 201–202  
*Joy v. North*, 520  
 judicial decision-making  
 anchoring effects and, 678–680  
 anti-inference bias and, 682  
 attitudes toward probabilistic evidence  
 and, 681–683  
 attitudinal models of, 664–665  
 base-rate neglect and, 682–683  
 Bayesian probability theory and, 3

- judicial decision-making (*cont.*)
- compromise effect and, 667, 670
  - contrast effect and, 667, 670
  - damage awards and, 675–677, 679–681
  - descriptive models of, 4–5, 9–10
  - ethnic bias and, 132
  - group decision-making dynamics and, 677, 688–690
  - group polarization and, 677, 689
  - hindsight bias and, 356–360, 673
  - implicit association tests (IATs) and, 684–685
  - inadmissible evidence and, 671–673
  - individual decision-making dynamics and, 677
  - judges *versus* laypersons and, 690–692
  - juries and, 666, 669, 677–681, 689, 692
  - law and economics research on, 666
  - mathematical models and, 9–10
  - moral heuristics and, 666
  - neo-institutionalist model of, 665
  - normative models of, 5–6, 8
  - numerical decisions and, 675–676
  - omission bias and, 673–675
  - “outrage model” and, 678
  - prejudice and, 683–686
  - prejudicial evidence and, 669
  - prescriptive models of, 6
  - priming and, 683–686
  - probable cause rulings and, 691
  - rational choice theory and, 664–666, 669
  - rules *versus* standards in, 686–688
  - stare decisis doctrine and, 674–675
  - status quo bias and, 674
  - story model of fact-finding and
    - coherence-based reasoning in, 667–669
    - sunk costs and, 675
- Kelo v. New London*, 383
- Kingston Fossil Plant (Tennessee), 763
- Kodak, 551–552
- Krell v. Henry*, 454–455
- lawyers’ fees, 275–278, 406
- Leegin* case, 558–559
- “Let’s Move” initiative, 741
- libertarian paternalism. *See also* paternalism
- decision architecture and, 23
  - environmental law and, 756–758, 769–771
  - externally imposed goals and, 183n12
  - goal-based decision-making and 176
  - nudges and, 726–727
  - organ donation and, 100
- litigation and settlement. *See also* plea bargaining
- adversarial posture of, 631–632
  - anchoring effects and, 627, 636
  - apologies and, 629, 631
  - availability heuristic and, 627
  - biased assimilation and, 633
  - compromise effect and, 634
  - confirmation bias and, 612–613
  - contrast effect and, 634
  - cost considerations and, 624–625, 629–630
  - fixed pie bias and, 632
  - framing effects and, 628
  - imperfect information and, 625
  - inaction inertia and, 630
  - information gathering and, 632–633
  - lawyers as agents in, 634–636
  - lawyers’ fee arrangements and, 634–635
  - litigant views of procedural justice and, 630–631, 636
  - loss aversion, 278–279, 628, 634–635
  - naïve realism and, 633
  - “negotiator’s dilemma” and, 631
  - nonmonetary influences and, 629–631
  - option devaluation and, 634
  - overoptimism bias and, 336, 341, 626
  - predictions and, 625–626
  - prospect theory and, 628
  - punishment magnitude gaps and, 632
  - rational choice theory and, 623–624
  - reactive devaluation and, 632, 636
  - regret aversion and, 629–630
  - risk tolerance and, 627–628
  - self-serving biases and, 626, 633, 636
  - settlement options and, 633–634
  - “shadow of the trial” and, 627
  - standard economic models of, 624–625
  - strategic bargaining decisions and, 625, 631
  - subadditivity and, 626
  - sunk costs and, 629
  - zone of possible agreement and, 625

- loss aversion
- affirmative action and, 284
  - asylum law and, 284
  - breach of contract and, 284
  - burden of proof rules, 280–281
  - civil and political rights (CPRs) and, 282–283
  - cognitive psychology and, 286–287
  - consumer behavior and, 274–275
  - contract law and, 275, 446–447, 454–455
  - default rules and, 279–280
  - deontological morality and, 283, 286–287
  - emotional aspects of, 271–272
  - endowment effect and, 149, 268, 270, 272–273, 275, 291, 312–317
  - environmental law and, 754, 764, 766–767
  - equity premium puzzle and, 271
  - evolutionary theory of common law and, 284–285
  - expected utility theory, 268–270, 273, 288–289
  - financial incentives for teachers and, 271
  - human rights and, 282–284
  - human welfare theories and, 288–289
  - information problems and, 277
  - lawmakers, among, 291
  - lawyers' contingency fees and, 275–278
  - legal norms' impact on reference points and, 269, 290–291
  - legal policymaking and, 289–290
  - litigation and settlement and, 278–279, 628, 634–635
  - normative implications, 288–291
  - nudges and, 723
  - omission bias and, 272, 275, 281, 292
  - plea bargaining and, 647–648
  - prediction of human behavior and, 269–270
  - prospect theory and, 268, 270–271, 273
  - regret and, 272, 274, 278–279
  - retirement savings and, 272
  - social and economic rights (SERs) and, 283–284
  - status quo bias, 268, 270–273, 275, 279, 289, 291–292
  - sunk costs and, 273, 275
  - taxation and, 600, 618
  - tort law and, 282
  - unjust enrichment and, 282
- mandatory minimum sentences, 653, 657, 660–661
- markets
- aggregation mechanisms and, 543–544
  - consumer behavior and, 542–543
  - firms' learning behavior and, 545–546
  - producer behavior and, 543–546
  - selection mechanisms and, 544–545
- Medicaid, 734
- Medicare, 103–104, 619, 734, 739
- moral attitudes. *See also* behavioral ethics;
- moral judgment
  - abortion and, 246
  - blood donation and, 250
  - breastfeeding and, 242
  - capital punishment and, 255
  - cigarette smoking and, 242, 248–252
  - date rape and, 256
  - dissensus and, 244, 246
  - energy use and, 252
  - gun control and, 243–245, 252–253
  - impact of source of legal regulations on, 253–255
  - legal regulation and, 244, 247–260
  - level of citizen involvement's impact on, 256
  - manner of legal regulation's impact on, 257–258
  - mechanisms of influence and, 247–258
  - nudges and, 248–250
  - perceived legitimacy of legal system and, 245–247
  - public information campaigns and, 241, 248, 252, 257
  - punishment's impact on, 245–246
  - recycling and, 248–252
  - regarding harms to third parties, 249, 255
  - school desegregation and, 254–255
  - self-interest and, 257, 259–260
  - sexual harassment and, 242–243
  - social meaning changes and, 250–253
  - U.S. Supreme Court's impact on, 258
- moral hazard, 494–496, 498, 502

- moral judgment. *See also* behavioral ethics;  
 moral attitudes  
 agent relativity and, 74  
 authority independence and, 62  
 belief and, 85–86  
 biases in helping others and, 78–80  
 blank slate model and, 61  
 charity heuristics and, 79–80  
 children’s cognitive development and,  
 61–63  
 deontology and, 75–78  
 distribution of goods and bads, 80–82  
 emotion’s role in, 76–78  
 equality heuristics and, 80–81  
 equity theory and, 63–64  
 genetics and, 62  
 history of research on, 61–63  
 identified victim cases and, 79  
 indirectness bias and, 73–74  
 intuition and, 64–65, 68–69  
 just world hypothesis and, 63  
 moralistic goals and, 70–71  
 morality of citizenship and, 82–84  
 naïve theories of citizenship and, 84  
 naturalism bias and, 75  
 normative theory of, 65–68  
 omission bias and, 68, 71–74  
 parochialism and nationalism in, 83–84  
 physical proximity and, 73  
 protected values and, 69–71, 73  
 psychological basis of biases in, 75–78  
 psychophysical numbing and, 79  
 trolley problem examples and, 71–73, 75  
 two-systems theory and, 75–76  
 utilitarianism and, 66–69, 71–72, 78
- multiattribute utility, 7
- MyData initiative (United Kingdom), 481
- myopia  
 consumer transactions and, 474, 482  
 retirement savings and, 103–105  
 taxation and, 603, 606–607, 613
- my-side bias, 20–21, 70
- naive theories, 22
- National Federation of Independent  
 Businesses v. Sebelius*, 254
- National Highway Transportation Safety  
 Administration (NHTSA), 758
- National Organ Transplantation Act of 1984,  
 99
- National School Lunch Act, 734
- naturalism bias, 75
- negative option marketing, 736
- Nelson v. James H. Knight DDS*, 242–243
- normality bias, 184
- normative model of decision-making, 6–8
- normative theory of decision-making, 3–6
- Nuclear Free Zones, 253–254
- nudges  
 active choice requirements and,  
 737–738  
 behavioral ethics and, 227–228  
 behavioral market failures and, 726  
 choice architecture and, 719  
 compliance without enforcement and,  
 723–724  
 default rules and, 721, 733–739  
 disclosure and, 722, 727–733  
 endowment effect and, 723  
 energy efficiency and, 721–722, 726–728,  
 736, 740, 742  
 environmental law and, 749, 756–758  
 food consumption and, 723, 725, 729–730,  
 735, 741  
 framing effects and, 722  
 hard shoves compared to, 256, 749  
 health care coverage and, 728–729, 734  
 informational cascades and, 723, 740  
 libertarian paternalism and, 726–727  
 loss aversion and, 723  
 market forces and, 726–727  
 as means to combat choice overload, 738  
 as means to combat obesity, 735  
 as means to combat procrastination and  
 inertia, 721–722, 735–736  
 moral attitudes and, 248–250  
 nutritional disclosure and, 728  
 optimism bias and, 724  
 probability assessment and, 724  
 retirement savings accounts and, 733–734  
 risk tolerance and, 724  
 salience of information and, 722–723,  
 739–740

- simplification of choices and, 719, 721, 726–727, 742  
social norms and, 723, 740–741
- Occupational Safety and Health Administration (OSHA), 733
- Office of Fair Trading, 508–509
- Office of Information and Regulatory Affairs, 483, 719, 757
- omission bias. *See also* status quo bias  
agent relativity and, 74  
evidence law and, 709  
judicial decision-making and, 673–675  
loss aversion and, 272, 275, 281, 292  
moral judgment and, 68, 71–74  
naturalism bias and, 75  
perceived causality and, 72  
physical proximity and, 73  
property law and, 378  
protected values and, 71, 73
- optimal contracting theory, 208–209
- Organization of Economic Cooperation and Development and Cooperation (OECD), 720, 751
- organ transplants, 98–100, 272
- outcome bias, 355–356, 366, 410–412
- overoptimism bias  
antitrust law and, 553, 557  
attribution bias and, 337–339, 524, 711, 713, 769–770  
automobile driving and, 147–148, 152, 335, 337, 339  
below-average effect and, 340  
benefits of, 340–341  
better-than-average effect and, 147  
blind spot bias and, 147–148, 152  
cigarette smoking and, 162–163, 341–342, 724  
consider-the-opposite strategies and, 341  
consumer behavior and, 336, 338, 340–341, 468, 474, 482  
contract law and, 445–446, 449, 452, 455  
at corporate law firms, 525, 548, 553  
criminal law and, 585–586  
empirical findings regarding, 336–340  
harm minimization and, 337  
health-related risk and, 341–342  
hindsight bias and, 362–363  
information levels and, 338  
insurance law and, 499, 504  
legal settlements and, 336, 341, 626  
marriage and, 342–343  
nudges and, 724  
plea bargaining and, 646–647  
representativeness bias and, 148, 163  
resiliency of, 339  
tort law and, 338, 413–414, 418–419  
warnings and prohibitions against, 249–250, 320, 721, 740
- paternalism  
hard paternalism and, 478  
libertarian paternalism and, 23, 100, 176, 183n12, 726, 756–758, 769–771  
soft paternalism and, 478
- Patient Protection and Affordable Care Act.  
*See* Affordable Care Act
- Pension Protection Act of 2006, 734
- Petition of Kinsman Transit Co.*, 411
- plea bargaining. *See also* litigation and settlement  
ambiguity aversion and, 345–346  
anchoring effects and, 657  
bounded rationality and, 645–646  
bounded willpower and, 648–649, 660  
criminal *versus* civil litigation and, 654–655  
discovery regimes and, 646, 654–657  
endowment effect and, 648, 659  
exculpatory evidence access and, 656  
fairness bias and, 649–650  
framing effects and, 658  
information deficits and, 645–646, 654–655  
in international context, 644  
lawyers' roles in, 659–660  
loss aversion and, 647–648  
mandatory minimum sentences and, 653, 657, 660–661  
minimization of probability-of-conviction uncertainty and, 654–656  
minimizing sentencing ambiguity and, 657–658

- plea bargaining (*cont.*)  
  opportunity costs and, 635, 645  
  overdiscounting and, 648–649  
  overoptimism bias and, 646–647  
  plea discounts *versus* trial penalties and,  
    651–654  
  pretrial detention's impact on, 658  
  prevalence of, 643  
  process costs' impact on, 658–659  
  prosecutors' preferences and, 650  
  prospect theory and, 647, 650  
  rational choice theory and, 643, 647–649,  
    657, 661  
  reverse proffers and, 656  
  risk aversion and, 646–648, 650, 658  
  self-serving bias and, 643, 646–647,  
    654–655  
  sentencing differentials and, 644, 652–654  
  three-strikes laws and, 653
- Polaroid, 526
- prescriptive theory of decision-making, 4–6
- prisoner's dilemma, 29, 39–40, 198
- property law  
  adverse possession and, 379, 387–389  
  bankruptcy proceedings and, 382  
  commercial property and, 383–384  
  crowding-out effect and, 378  
  development rights and, 384–386  
  eminent domain and, 378, 381  
  endowment effect and, 384–385  
  existing *versus* prospective use and,  
    384, 387  
  fungible property and, 380–387  
  home-protective laws and, 383  
  in-kind *versus* monetary remedies and,  
    378, 386, 394, 397–398  
  nonfungible property and, 382, 385, 387, 389  
  omission bias and, 378  
  ownership *versus* possession and, 387–389  
  personhood theory and, 382, 399  
  physical *versus* nonphysical injuries to land  
    and, 384–385  
  “property as bundle of sticks” definition  
    and, 380–381  
  “property as thing” definition, 380–381  
  property rules *versus* liability rules and,  
    390–395  
  redistribution and, 395–397  
  regret avoidance bias and, 378  
  self-development and, 386, 394  
  tenancy and, 379, 387–388, 395–397  
  transactions costs and, 390–391  
  zoning regulations and, 38, 384–385
- prosocial behavior  
  altruism and, 199–200  
  contract law and, 206–207  
  corporate and securities law and, 207–209  
  “crowding out phenomenon” and, 204, 209  
  dictator game's measurement of, 198–200  
  environmental preservation and, 203–204  
  fiduciary duties and, 208  
  Jekyll-Hyde syndrome and, 201–202  
  legal implications of, 202–209  
  negligence and, 204–206  
  optimal contracting and, 208–209  
  pervasiveness of, 198  
  social cues and, 201  
  social dilemma game's measurement of,  
    198, 200–201  
  social norms and, 203  
  spite and, 199–200  
  as supplement or replacement to  
    regulation, 203–204  
  tort liability and, 205  
  ultimatum game's measurement of,  
    198–200
- prospect theory  
  criminal law and, 579–581  
  environmental law and, 760  
  loss aversion and, 268, 270–271, 273  
  plea bargaining and, 647, 650  
  risky choices and, 176  
  tort law and, 413
- protected values, 69–71, 73
- Psychology, Public Policy & Law* (journal), 126
- public goods games  
  altruistic punishment and, 35, 47  
  antisocial punishment and, 51–54  
  basic features of, 34  
  collective action problems and, 45  
  conditional cooperation and, 42–45, 54  
  cooperation and, 30, 34  
  efficiency seeking and, 35–36  
  external punishment and, 39



- free riding and, 34–40, 42, 44–51, 54  
 guilt aversion and, 39, 43, 54  
 imperfect observation and errors in, 49  
 indirect reciprocity and, 50  
 inequality aversion and, 35–37, 43  
 institutionalized punishment and, 49–50  
 intentions' impact in, 35–37  
 internal punishment and, 39  
 linear version of, 34  
 moral judgments in, 38–40  
 partners condition and, 46  
 perfect stranger condition and, 47  
 prisoner's dilemma and, 39–40  
 punishment and, 35–37, 39–40, 45–54  
 reciprocity and, 30, 33–37, 40–43, 46, 48, 50–51  
 reputation and, 50–51  
 rewards and, 50  
 selective incentives and, 45  
 self-interest and, 30  
 stranger condition and, 46–47
- rational choice theory. *See also homo economicus*  
 antitrust law and, 540–541  
 behavioral law and economics' challenges to, 93, 95–97, 101–102, 111–114, 116, 143, 164, 167–172, 175, xi  
 contract law and, 439, 441–446  
 corporate and securities law and, 518–521, 523–524, 526–527, 534  
 corporate firms and, 546–550  
 criminal law and, 571–574, 589  
 environmental law and, 759, 767  
 judicial decision-making and, 664–666, 669  
 litigation and settlement and, 623–624  
 plea bargaining and, 643, 647–649, 657, 661  
 tautological problems with, 117  
 taxation and, 599–601, 604–605  
 utility maximization and, 68, 143–145, 167–169, 175, 195–198, 200–201, 209, 268, 278, 288, 420n11, 439, 441, 460, 497, 518, 523, 623–624, 649, 665, xi  
 weakness of will and, 116  
 welfare economics and, 610–611, 614, 767, 773, xi
- recency effect, 156  
 reciprocity  
   efficiency seeking and, 36–37  
   heuristic of, 14  
   indirect reciprocity and, 50  
   inequality aversion and, 35–37, 43  
   intentions and, 36–37  
   norms of, 724  
   strong negative forms of, 30, 33–36, 46, 48, 51  
   strong positive forms of, 30, 33–34, 40–42, 51  
 recycling, 248–252  
 resale price maintenance (RPM), 557–559  
 Resource Conservation and Recovery Act (RCRA), 763  
*Restatement First of Contracts*, 441  
*Restatement of Torts*, 368  
*Restatement Second of Contracts*, 441, 450, 460  
 retirement savings  
   default rules and, 733–734  
   loss aversion and, 272  
   myopia and, 103–105  
   nudges and, 733–734
- Sarbanes-Oxley Act, 151n19, 157, 533  
 Save More Tomorrow program, 102  
 second-order thoughts, 118  
 Securities and Exchange Commission, 533  
 self-serving bias. *See also* overoptimism bias  
   litigation and settlement and, 626, 633, 636  
   plea bargaining and, 643, 646–647, 654–655  
 set point of individual happiness, 101–102  
 sexual harassment, 242–243  
 Sherman Act, 540, 551  
 SMarT savings program, 179  
 smoking. *See* cigarette smoking  
 Smoking Prevention and Tobacco Control Act, 731  
 social dilemma game, 198, 200–201  
 social norms  
   behavioral ethics and, 224–225  
   distracted driving and, 731, 741  
   environmental law and, 757–758, 775  
   nudges and, 723, 740–741  
   prosocial behavior and, 203  
   seat belts and, 723–724

- social order. *See also* public goods games  
 behavior of other people and, 40–45  
 internalized norms and, 37–40, 53  
 personal ethics and, 30–31  
 prisoner's dilemma and, 29  
 punishment and, 31, 45–51  
 rule of law and, 28–29, 51–53  
 self-enforcement and, 29–30  
 self-interest and, 28–29, 54
- Social Security, 103–104, 616, 619
- status quo bias. *See also* omission bias  
 environmental law and, 753–754, 757,  
 765–767  
 judicial decision-making and, 674  
 loss aversion, 268, 270–273, 275, 279, 289,  
 291–292  
 taxation and, 600, 606, 615, 618
- sunk-cost effect, 15, 273, 275, 629, 675
- Superfund sites, 348, 764
- Swann v. Charlotte-Mecklenburg Board of  
 Education*, 254
- System 1 thinking  
 behavioral ethics and, 214–215, 218–219,  
 225–227, 231  
 defining features of, 18–19, 171  
 moral judgment and, 75–77
- System 2 thinking  
 behavioral ethics and, 215–216, 219–220,  
 227, 230–231  
 defining features of, 18, 171  
 moral judgment and, 75–77
- taxation  
 advanced payment and, 581  
 allocative function of, 610–611  
 ambiguity aversion and, 346  
 behavioral modification and, 605–607  
 child tax credits and, 609  
 communication strategies and, 614  
 compliance matters and, 603–605  
 differential taxation and, 83  
 disaggregation bias and, 609–610, 614  
 distributive function of, 610–611  
 endowment effect and, 600–601, 618  
 fiscal balance and, 617–619  
 fiscal cliff negotiations (2012) and, 615–617  
 flat taxes and, 83, 608–609  
 form aversion and, 610  
 framing effects and, 599–600, 609, 614  
 government enforcement and, 604  
 loss aversion and, 600, 618  
 market salience of taxes and, 602–603, 612  
 metric effect and, 608  
 myopia and, 603, 606–607, 613  
 normative analysis of, 610–615  
 perceptions of fairness and, 604  
 political salience of taxes and, 602–603, 611  
 privatization effect and, 614  
 progressivity and, 608–610, 614–615  
 rational choice theory and, 599–601,  
 604–605  
 retirement savings plans and, 606–607  
 sin taxes and, 605  
 status quo bias and, 600, 606, 615, 618  
 subadditivity and, 609  
 substitution effect and, 612–613  
 tax evasion and, 604  
 tax system design and, 607–617
- A Theory of Justice* (Rawls), 65
- thick rationality, 146–147
- Thinking, Fast and Slow* (Kahneman), 18, 215
- thin rationality, 146–147, 521
- three-strikes laws, 653
- TimeWarner, 526
- Tire Fuel Efficiency Consumer Information  
 Program, 758
- tort law  
*versus* administrative regulation, 417–420  
 affective forecasting and, 107–108  
 aggregate social welfare goal of, 407–408  
 aggregative cost-benefit test and, 430  
 alternative-care settings and, 408–409, 418  
 attorneys' contingency fees and, 406  
 availability heuristic and, 414, 418–419  
 bellwether trials and, 432  
 bilateral-care settings and, 408–409, 413, 415  
 comparative negligence rules and, 409,  
 416, 431–432  
 contributory negligence rules and,  
 409–410, 413, 415–416  
 corrective justice norm and, 183  
 deterrence and, 81, 409, 413–415, 418–420,  
 423–424, 432

diffuse liability and, 428  
 employees *versus* independent contractors  
 and, 426–428  
 environmental disasters and, 431  
 fixed accident losses and, 413  
 framing effects and, 420–421  
 Golden Rule and, 421  
 Hand Formula and, 406, 411  
 hedonic adaptation and, 421–423  
 heterogeneity of victims and, 428–432  
 hindsight bias and, 356–357, 410–412, 414, 419  
 indirectness effect and, 427  
 insurance and, 410, 417–418, 425  
 jurors and, 412, 420–421, 423–425  
 law and economics research and, 405, 410  
*versus* law of unjust enrichment, 269, 282  
 legislated damages caps and, 205, 680  
 market share liability and, 432  
 negligence and, 406–418, 424, 426, 430–431  
 neoclassical economic model for accidents  
 and, 407–408, 415–417  
 no liability regimes and, 409, 415–416  
 optimal enforcement and, 424  
 outcome bias and, 410–412  
 outgroup homogeneity bias and, 428–432  
 overoptimism bias and, 338, 413–414,  
 418–419  
 pain and suffering damages and, 420  
*versus* Pigovian taxes, 417  
 probabilistic nature of, 405–406, 409–414,  
 417–420  
 probability insensitivity bias and, 413  
 probability of successful lawsuit and, 424  
 prospect theory and, 413  
 punitive damages and, 81, 423–425  
 pure negligence rules and, 409, 415  
 remedies and, 406  
 respondent superior doctrine and, 426  
 risk-based regimes *versus* harm-based  
 regimes, 419–420  
 risk of insolvency's impact on, 415–419  
 safe harbors and, 368  
 split-award statutes and, 138  
 standard of care with multiple victims and,  
 430–432  
 strict liability regimes and, 361, 407,  
 409–416, 418

underestimation of group variability and,  
 428–429  
 underresearched areas in, 425–432  
 variable accident losses and, 413  
 vicarious liability and, 425–426  
 Treasury Department, 731, 734  
 trolley problem, 71–73, 75  
 trust game, 33–34  
 Truth-in-Lending Act, 485

ultimatum game  
 basic rules in, 32  
 departure from *homo economicus*  
 assumptions in, 33  
 equity theory and, 64  
 fairness bias and, 649  
 heterogeneity of outcomes in, 178  
 inequality aversion and, 36  
 negative reciprocity and, 34  
 prosocial behavior and, 198–200  
 transfers of entitlements and, 391–392

Uniform Anatomical Gift Act, 100  
 unilateral change clause, 473–474  
 United Kingdom, 100, 481, 483, 501, 505, 508,  
 720, 737  
*United States v. Bagley*, 656  
*United States v. Carroll Towing Co.*, 328, 406n2  
*United States v. Conroy*, 656  
*United States v. Rose Marks et al.*, 493n1  
 unjust enrichment, 282  
 utilitarianism. *See also* consequentialism  
 altruism and, 84  
 conflicts with intuition and, 68–69  
 deontology and, 75  
 expected utility theory and, 68  
 motivation and, 80n7  
 as a normative theory of moral judgment,  
 66–69, 71–72, 78  
 people as units of analysis in, 7  
 person's identity as irrelevant in, 8, 83  
 utility maximization. *See under* rational  
 choice theory

verbatim memories, 173–174  
 victim impact statements, 102  
 Vietnam War, 259

- Walmart, 486, 741  
*The Walt Disney Co. Derivative Litigation*  
2006, 368, 529  
welfare economics, 68, 610–611, 614, 767, 773, xi  
*Weyerhaeuser Co. v. Ross-Simmons*  
*Hardwood Lumber Co.*, 541  
wicked environments, 175  
*Williams v. Walker-Thomas Furniture Co.*,  
455–456  
willingness to accept (WTA). *See under*  
endowment effect  
willingness to pay (WTP). *See under*  
endowment effect  
World Trade Organization (WTO), 751



