ADVANCES IN Behavioral Economics

Edited by

COLIN F. CAMERER

GEORGE LOEWENSTEIN

AND MATTHEW RABIN

Advances in Behavioral Economics

The Roundtable Series in Behavioral Economics

The Roundtable Series in Behavioral Economics aims to advance research in the new interdisciplinary field of behavioral economics. Behavioral economics uses facts, models, and methods from neighboring sciences to establish descriptively accurate findings about human cognitive ability and social interaction and to explore the implications of these findings for economic behavior. The most fertile neighboring science in recent decades has been psychology, but sociology, anthropology, biology, and other fields can usefully influence economics as well. The Roundtable Series publishes books in economics that are deeply rooted in empirical findings or methods from one or more neighboring sciences and advance economics on its own terms—generating theoretical insights, making more accurate predictions of field phenomena, and suggesting better policy.

Colin Camerer and Ernst Fehr, Series Editors

The Behavioral Economics Roundtable

Henry Aaron
George Akerlof
Linda Babcock
Colin Camerer
Peter Diamond
Jon Elster
Ernst Fehr
Daniel Kahneman
David Laibson
George Loewenstein
Sendhil Mullainathan
Matthew Rabin
Thomas Schelling
Eldar Shafir
Robert Shiller
Cass Sunstein
Richard Thaler
Richard Zeckhauser

Advances in Behavioral Economics

Edited by
COLIN F. CAMERER, GEORGE LOEWENSTEIN,
and MATTHEW RABIN

RUSSELL SAGE FOUNDATION, NEW YORK, NEW YORK

PRINCETON UNIVERSITY PRESS

PRINCETON AND OXFORD

Published by Princeton University Press,
41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press,
3 Market Place, Woodstock, Oxfordshire OX20 1SY

and Russell Sage Foundation,
112 East 64th Street, New York, New York 10021

Library of Congress Cataloging-in-Publication Data
Advances in behavioral economics / edited by Colin F. Camerer, George Loewenstein, and
Matthew Rabin.
p. cm. — (The roundtable series in behavioral economics)
Includes bibliographical references and index.
ISBN 0-691-11681-4 (alk. paper) — ISBN 0-691-11682-2 (pbk.: alk. paper)
1. Economics—Psychological aspects. I. Camerer, Colin, 1959– II. Loewenstein, George.
III. Rabin, Matthew, 1963– IV. Series.

HB74.P8A375 2003
330′.01′9—dc21 2003044481

This book was composed in Times

Printed on acid-free paper. ∞

www.pup.princeton.edu
www.russellsage.org

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

To Daniel Kahneman, Richard Thaler, Amos Tversky, *and* Eric Wanner

C O N T E N T S

C O N T R I B U T O R S

George A. Akerlof is the Koshland Professor of Economics at the University of California, Berkeley. He received his Ph.D. from MIT in 1966, at which time he joined the faculty at Berkeley. In 2001 he was corecipient of the Nobel Prize in Economics for his work on the role of asymmetric information in markets. He has also pioneered the application of sociology and psychology to the workings of the macroeconomy. He has proposed efficiency wage explanations for unemployment. According to these explanations, employers, because of concerns about worker morale, may not wish to reduce wages to market clearing. He has also explored reasons why firms might be slow to change wages and prices, thereby explaining the business cycle and the effectiveness of monetary policy. Akerlof has been vice president and member of the executive committee of the American Economics Association.

Linda Babcock is the James Mellon Walton Professor of Economics at the Heinz School of Public Policy and Management at Carnegie Mellon University. Babcock earned a Ph.D. in economics from the University of Wisconsin and has received numerous research grants from the National Science Foundation. She teaches negotiation and has won the school's highest teaching award twice. She has investigated how cognitive biases in negotiator beliefs cause conflict in negotiations, as well as the effect of various tort reforms on negotiation impasses, and the role of social comparisons in affecting negotiated outcomes. Her research has appeared in the most prestigious economics, industrial relations, and law journals. Her most recent research examines the situational factors that affect gender differences in negotiation and is summarized in her recent book, *Women Don't Ask: Negotiation and the Gender Divide* (Princeton, 2003).

Shlomo Benartzi is an associate professor at UCLA's Anderson Graduate School of Management. Benartzi received his Ph.D. from Cornell University's Johnson Graduate School of Management. His research investigates participant behavior in defined contribution plans. In particular, his current work examines how participants make investment choices in retirement saving plans and how employee saving rates could be increased. Benartzi's work has been published in the *Journal of Political Economy, American Economic Review, Journal of Finance, and Management Science.* His work been discussed in the *Economist, Financial Times, Investor's Business Daily,* the *Los Angeles Times, Money Magazine,* the *New York Times, Plan Sponsor, Pensions and Investments*, the *Wall Street Journal*, and CNBC. Benartzi served on the ERISA Advisory Council of the U.S. Department of Labor, and he currently serves on the advisory board of Morningstar and the Investment Advisory Council of the Alaska State Pension.

Colin F. Camerer is the Axline Professor of Business Economics at Caltech, in Pasadena, California, where he teaches both psychology and economics. Camerer earned a Ph.D. in behavioral decision theory in 1981 from the University of Chicago, and worked at Kellogg, Wharton, and Chicago business schools before Caltech. His research in behavioral economics focuses mostly on theories of risky decision making and strategic behavior in games. He has also done experiments on price bubbles and "cascades" in asset markets, creation of organizational culture in the form of "codes," and is now doing neuroscientific imaging experiments on behavior in games. Camerer has also analyzed field data on hot-hand biases and commitment escalation in NBA basketball, and the labor supply of New York City cab drivers. Besides nearly 100 journal articles and book chapters, he is the coauthor or editor of four books, and the author of *Behavioral Game Theory* (Princeton, 2003). Camerer was the first behavioral economist to become a Fellow of the Econometric Society, in 1999, and was president of the Economic Science Association 2001–03.

VINCENT CRAWFORD earned an A.B. summa cum laude in Economics from Princeton in 1972, and a Ph.D. in economics from the Massachusetts Institute of Technology in 1976. Since 1976 he has worked at the University of California, San Diego, where he is now Professor of Economics. He has held visiting positions at Harvard, Princeton, Australian National University, University of Canterbury, and the Ecole des Hautes Etudes en Sciences Sociales. Honors include election as Fellow of the Econometric Society, a Guggenheim fellowship, election to the Council of the Game Theory Society and to an Overseas Fellowship at Churchill College, Cambridge, and several invited lectures. His work focuses on game theory and its applications, from early work on learning in games, bargaining and arbitration, matching markets, coordination, and strategic communication to recent work interpreting the results of experiments and conducting experiments to study players' mental models of other players.

PETER DIAMOND is an Institute Professor at the Massachusetts Institute of Technology, where he has taught since 1966. He received his B.A. in Mathematics from Yale University in 1960 and his Ph.D. in Economics from MIT in 1963. He has been president of the Econometric Society and is president of the American Economic Association. He is a founding member of the National Academy of Social Insurance, where he has been president and chair of the board. He is a Fellow of the American Academy of Arts and Sciences and a Member of the National Academy of Sciences. He was the recipient of the 1980 Mahalanobis Memorial Award and the 1994 Nemmers Prize. He has written on behavioral economics, public finance, social insurance, uncertainty and search theories, and macroeconomics. His writings on social security reflect his awareness of the importance of behavioral issues.

ERNST FEHR is a professor in Microeconomics and Experimental Economics at the University of Zürich. He is director of the Institute for Empirical Research in Economics at the University of Zürich and of the Ludwig Boltzmann Institute for the Analysis of Economic Growth in Vienna. Ernst Fehr graduated at the University of Vienna in 1980, where, in 1986, he also earned his doctorate. His research focuses on the proximate patterns and the evolutionary origins of human altruism and the interplay between social preferences, social norms, and strategic interactions. He has conducted extensive research on the impact of social preferences on competition, cooperation, and on the psychological foundations of incentives. More recently he has worked on the role of bounded rationality in strategic interactions. He is on the editorial board of the *Quarterly Journal of Economics*, the *European Economic Review*, *Games and Economic Behavior*, the *Journal of the European Economic Association*, the *Journal of Public Economics and Experimental Economics*. He won the Gossen Price of the German Economic Association in 1999 and the Hicks-Tinbergen Medal of the European Economic Association in 2000. He has given several keynote lectures, among them the Frank Hahn Lecture at the annual Congress of the Royal Economic Society 2001, the Schumpeter Lecture at the annual Congress of the European Economic Association 2001, and an invited Lecture at the Eighth World Congress of the Econometric Society in 2000. He is president of the Economic Science Association for the years 2003–5.

ROBERT H. FRANK is the H. J. Louis Professor of Economics at Cornell's Johnson Graduate School of Management. He received his B.S. in mathematics from Georgia Tech in 1966, then taught math and science for two years as a Peace Corps Volunteer in rural Nepal. He received his M.A. in statistics from the University of California, Berkeley, in 1971 and his Ph.D. in economics in 1972, also from UC Berkeley. During leaves of absence from Cornell, Frank was chief economist for the Civil Aeronautics Board from 1978 to 1980, a Fellow at the Center for Advanced Study in the Behavioral Sciences in 1992–93, and a professor of American Civilization at the Ecole des Hautes Etudes en Sciences Sociales in Paris in 2000–1. Frank's books, which include *Choosing the Right Pond, Passions Within Reason, Microeconomics and Behavior, Principles of Economics* (with Ben Bernanke), and *Luxury Fever*, have been translated into nine languages. *The Winner-Take-All Society*, coauthored with Philip Cook, received a Critic's Choice Award, was named a Notable Book of the Year by the *New York Times*, and was included in *Business Week*'s list of the ten best books of 1995.

SHANE FREDERICK is an assistant professor of management science at the Massachusetts Institute of Technology.

SIMON GÄCHTER is a professor of Economics at the University of St. Gallen. He teaches courses on microeconomics, game theory, organizational and labor economics, experimental economics, and economics and psychology. Gächter received his Ph.D. in Economics in 1994 at the University of Vienna. After postgraduate lecturer positions at the universities of Vienna and Linz, Gächter became an assistant professor at the University of Zürich. In 2000 he became a full professor of Economics at the University of St. Gallen. His main research interests and publications are on behavioral issues of voluntary cooperation and punishment, wage formation, and incentive contracting. Gächter is affiliated with the MacArthur Foundation research network on social norms and preferences and the CESifo research network on Employment and Social Protection.

DAVID GENESOVE is currently an associate professor of Economics at the Hebrew University of Jerusalem. He earned his Ph.D. at Princeton University in 1991, and taught at the Massachusetts Institute of Technology from 199l to 1998. He has been an editor of the *Journal of Industrial Economics* since 1998. Genesove has written extensively on industrial organization, producing empirical studies on a wide variety of markets, including those for used cars, fish, housing, sugar, and daily newspapers.

Itzhak Gilboa is a professor at Eitan Berglas School of Economics and Recanati School of Business, Tel Aviv University, and a fellow of Cowles Foundation for Research in Economics, Yale University. He graduated from Tel Aviv University (in economics) in 1987 and was on the faculty of the Kellogg School of Management, Northwestern University, for ten years before returning to Israel. His main topic of research is decision under uncertainty in situations where there is too little information for the generation of a Bayesian prior. Together with David Schmeidler, Gilboa has developed axiomatic theories of decision making when information is modeled by sets of prior probabilities and by cases. Their joint project may be viewed as providing decision theories and axiomatic foundations for formal models representing information and belief that differ from the Bayesian one. The emphasis of this project is on scarcity of information rather than on irrational behavior of mistakes. Other topics that Gilboa has worked on include game theory, computational complexity, social choice, and consumer behavior.

Uri Gneezy is an associate professor of Behavioral Science at the University of Chicago Graduate School of Business, where he teaches negotiation. Gneezy earned a Ph.D. in economics in 1997 from the Center of Economic Research at Tilburg University, and worked at Haifa University and the Technion in Israel before Chicago. His research in behavioral economics investigates the effect of incentives on behavior in labor markets and its relation to sociological factors such as ethnicity and gender. Other areas of research are behavioral finance and behavioral game theory. The work is based mainly on laboratory experiments and field studies.

Robert M. Hutchens is a Professor in the Department of Labor Economics at Cornell's School of Industrial and Labor Relations. His early research dealt with the economics of government transfer programs and his later research has concentrated on long-term implicit contracts and on employer policy toward older workers. Hutchens has served as a policy fellow at the Brookings Institution, associate editor at the *Industrial and Labor Relations Review*, chairman of the Department of Labor Economics at Cornell, visitor at the University of British Columbia, and as a research fellow at the Institute for the Study of Labor (IZA).

Daniel Kahneman, winner of the 2002 Nobel Prize in Economics is currently a professor of Psychology and Public Policy at Princeton University. Formerly a professor of psychology at the University of California, Berkeley, a fellow at the Canadian Institute for Advanced Research, a professor of Psychology at the University of British Columbia, a fellow at the Center for Advanced Study in the Behavioral Sciences, and a professor at the Hebrew University in Jerusalem, Kahneman is a member of the American Academy of Arts and Sciences and the National Academy of Sciences. He is a fellow of the American Psychological

Association, the American Psychological Society, the Society of Experimental Psychologists, and the Econometric Society. He has been the recipient of numerous awards, among them the Distinguished Scientific Contribution Award of the American Psychological Association, the Warren Medal of the Society of Experimental Psychologists, and the Hilgard Award for Career Contributions to General Psychology. He earned a Ph.D. at the University of California, Berkeley.

JACK L. KNETSCH is a professor emeritus at Simon Fraser University in British Columbia, where he has taught and conducted research in the areas of behavioral economics, environmental economics, and law and economics for the past thirty years. He holds degrees in Soil Science, Agricultural Economics, Public Administration, as well as a Ph.D. in Economics from Harvard University. He has been with private and public agencies and organizations in the United States and Malaysia, and was at George Washington University before moving to Simon Fraser University. He has accepted visiting appointments at universities in Europe, Asia, Australia, as well as North American. Most of his behavioral economics research has involved tests of the disparity in people's valuations of gains and losses, and the implications of the observed differences in various areas of economic and policy interest. More recent work has included research on time preferences and measures of welfare change.

DAVID LAIBSON holds a B.A. from Harvard University, an M.Sc. from the London School of Economics and a Ph.D. from the Massachusetts Institute of Technology. In1994 Laibson joined the economics faculty at Harvard University, where he is currently a professor of Economics. Laibson is a member of the National Bureau of Economic Research, where he is a research associate in the Asset Pricing, Economic Fluctuations, and Aging Working Groups. Laibson has received a Marshall Scholarship and grants from the National Science Foundation, the MacArthur Foundation, the National Institute on Health, the Sloan Foundation, and the John M. Olin Foundation. In 1999 he received the Phi Beta Kappa Prize for Excellence in Teaching. Laibson's research focuses on the topic of psychology and economics. He is currently working in the fields of macroeconomics, intertemporal choice, decision and cognitive sciences, behavioral finance, and experimental economics.

GEORGE LOEWENSTEIN is a professor of Economics and Psychology at Carnegie Mellon University. He received his Ph.D. from Yale University in 1985 and since men has held academic positions at the University of Chicago and Carnegie Mellon University, and fellowships at the Center for Advanced Study in the Behavioral Sciences, the Institute for Advanced Study in Princeton, the Russell Sage Foundation, and the Institute for Advanced Study in Berlin. His research focuses on applications of psychology to economics, and his specific interests include decision making over tune, bargaining and negotiations, psychology and health, law and economics, the psychology of adaptation, the psychology of curiosity, and "out of control" behaviors such as impulsive violent crime and drug addiction.

CHRISTOPHER MAYER is an associate professor of Real Estate at Wharton School, University of Pennsylvania. Mayer, a real estate expert, has earned widespread recognition for his teaching and publications in his field. His research explores a wide variety of topics, including the implications of behavior economics for the cyclical nature of real estate, both in housing and commercial real estate markets. Mayer has also written on the link between the housing market and local school spending, and the impact of taxes, land-use regulations, and pollution on housing and stock market values. He is continuing a long-term project on the airline industry, examining scheduling practices and congestion. Mayer has authored numerous academic articles on these subjects, and he is frequently interviewed in the national media, including the *Wall Street Journal*, CNBC, the *Washington Post*, and the *New York Times*. Mayer holds a B.A. in Math and Economics from the University of Rochester and a Ph.D. in Economics from MIT. He has previously held positions at Columbia University, the University of Michigan, and the Federal Reserve Bank of Boston.

TERRANCE ODEAN is an associate professor of Finance at the Haas School of Business at the University of California, Berkeley. He earned, a B.A. in Statistics at UC Berkeley in 1990 and a Ph.D. in Finance from the university's Haas School of Business in 1997. He taught finance at UC Davis from 1997 through 2001. As an undergraduate at Berkeley, Odean studied Judgment and Decision Making with Daniel Kahneman. This led to his current research focus on how psychologically motivated decisions affect investor welfare and securities prices. During the summer of 1970, he drove a yellow cab in New York City.

TED O'DONOGHUE is an assistant professor of Economics at Cornell University. He earned a Ph.D. in Economics from University of California, Berkeley, in 1996, and spent one year as a postdoctoral fellow in the Center for Mathematical Studies in Economics and Management Sciences at Northwestern University before joining the Economics Department at Cornell. O'Donoghue's research in behavioral economics has been primarily on the topic of intertemporal choice. He has investigated the role that self-control problems might play in procrastination, addiction, (not) planning for retirement, and risky behavior among youths. He has also studied the implications of mispredictions of future utility.

MATTHEW RABIN is a professor of Economics at the University of California, Berkeley. He earned his B.S. in Mathematics and in Economics from the University of Wisconsin–Madison in 1984, and his Ph.D. in Economics from MIT in 1989. His research includes developing formal theoretical models of fairness and risk preferences, biases in predicting preferences, cognitive biases and inferential errors, and procrastination and other forms of self-control problems. He is a fellow of the Econometric Society, the American Academy of Arts and Sciences, and the MacArthur Foundation, and he was awarded the John Bates Clark Medal by the American Economic Association in 2001.

ALDO RUSTICHINI is a professor of Economics at the University of Minnesota. He has degrees in Philosophy, Economics, and Mathematics. His main activity has been in different branches: general equilibrium, growth theory, political theory, auction theory, decision theory, experimental economics and neuroscience. His contributions include precise estimates of the rate of convergence to truth-telling equilibria in auctions, the importance of indeterminacy in dynamic general equilibrium models, the detrimental effect of social groups in growth, and the existence (and nonexistence) of competitive equilibria in economies with private information. In decision theory, Rustichini has developed a formal theory of unawareness, and an axiomatic theory of preference for flexibility with applications to temptation and self-control. He has done research in experimental economics: he has with Uri Gneezy started the analysis of the paradoxical effects of rewards and punishments. He has determined significant differences in the competitive behavior of women and men. He has analyzed the effects of moods and emotions on cooperative behavior. Rustichini has in the last years focused on the analysis of the brain as a Bayesian, optimizing, decision machine. He is associate editor of the *Journal of Economic Theory, Journal of Mathematical Economics, Review of Economic Dynamics, and Games and Economic Behavior.*

DAVID SCHMEIDLER's research in recent years has dealt mainly with the informational aspects of decisions under uncertainty and belief representations. His other works are in the fields of cooperative and noncooperative games, classical functional analysis, and microeconomics. The latter includes works on topics of general equilibrium, implementation, and equity. He divides his time as professor at Tel Aviv University between Mathematics and Economics: specifically, he is affiliated with the Department of Statistics and Operations Research at the School of Mathematical Sciences, as well as the Faculty of Management. He is also a professor in the Department of Economics at Ohio State University. He wrote his Ph.D. thesis at the Institute of Mathematics of the Hebrew University in Jerusalem, Under the supervision of R. J. Aumann. It dealt with cooperative and noncooperative games and with general equilibrium.

KLAUS M. SCHMIDT has been professor of Economics at the University of Munich since 1995. He studied Economics and Political Science and completed his Ph.D. in Economics in a joint program of the University of Bonn and the London School of Economics in 1991. In 1995, he earned his Habilitation at the University of Bonn. He taught as a visiting professor at MIT and Stanford University. His research focuses on game theory, contract theory, and behavioral economics. In particular, he is interested in the impact of fairness and reciprocity on human behavior and on the optimal design of contracts and institutions. Schmidt serves as editor of the *European Economic Review* and as associate editor of the *Review of Economic Studies* and the *RAND Journal of Economics*. In 2001 he was awarded the Gossen-Prize of the German Economic Association and the Research Prize of the Berlin-Brandenburg Academy of Sciences.

ELDAR SHAFIR is a professor of Psychology and Public Affairs in the Department of Psychology and the Woodrow Wilson School of Public Affairs at Princeton University. He received his Ph.D. in Cognitive Science from the Massachusetts Institute of Technology in 1988, and was a postdoctoral scholar at Stanford University. He has held visiting positions at the University of Chicago Graduate

School of Business, the Kennedy School of Government, the Institute for Advanced Studies of the Hebrew University, and the Russell Sage Foundation. His research focuses on descriptive studies of decision making and their implications for economics and rationality. He received the Hillel Einhorn New Investigator Award from the Society for Judgment and Decision Making, and the Chase Memorial Award.

HERSH M. SHEFRIN is the Mario L. Belotti Professor of Finance at Santa Clara University. Shefrin earned his Ph.D. at the London School of Economics in 1974. Before joining Santa Clara, he taught at the University of Rochester. His work in behavioral economics and finance focuses on the manner in which self-control, prospect theory, regret, and heuristics impact financial decisions and financial judgments. In the 1980s, he focused on the impact of behavioral concepts on household savings behavior, the disposition effect (a term he coined to describe the disposition of investors to sell winners too early and hold losers too long), and the attractiveness of cash dividends to investors, despite tax disadvantages. In the 1990s he worked to develop behavioral theories of portfolio selection, asset pricing theory, and ethics. His work on behavioral portfolio theory was accorded the William F. Sharpe Award in 2000, and his work in behavioral ethics was accorded a Graham and Dodd Scroll in 1993. Shefrin's book *Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing* (Harvard Business School Press, 1999, Oxford University Press, 2002) is the first comprehensive treatment of behavioral finance, written for both business students and financial practitioners. He edited a the three-volume collection, *Behavioral Finance* (Edward Elgar, 2002).

CHRIS STARMER is a professor of Experimental Economics at the University of Nottingham. Starmer was awarded a Ph.D. for an experimental investigation of decision under risk in 1992 from the University of East Anglia (UEA). He worked as a lecturer then senior lecturer at UEA and was visiting associate professor at Caltech before moving to Nottingham in 2000. His research in behavioral economics investigates decision making under risk, equilibrium selection in games, and dynamic decision making. One stream of this work with a public policy focus has involved appraising and developing approaches to the valuation of nonmarketed goods. He has published articles on these topics in *American Economic Review, Econometrica, Economic Journal, Economica, Journal of Economic Literature, Quarterly Journal of Economics*, and *Review of Economic Studies.* Starmer is currently Director of the Centre for Decision Research and Experimental Economics (CeDEx) at the University of Nottingham.

RICHARD H. THALER is the Robert P. Gwinn Professor of Economics, Finance, and Behavioral Science at the University of Chicago's Graduate School of Business, where he is the director of the Center for Decision Research. He is also a research associate at the National Bureau of Economic Research, where he codirects the behavioral economics project. Thaler is considered one of the pioneers in the attempt to fill the gap between psychology and economics. Among the problems he has worked on are self control, savings, mental accounting, fairness, the endowment effect, and behavioral finance. He is the author of the books *The Winner's Curse* and *Quasi Rational Economics*, and is an editor of the collection *Advances in Behavioral Finance.* He writes a series of articles in the *Journal of Economics Perspectives* under the heading "Anomalies."

The late AMOS TVERSKY earned his Ph.D. in Psychology from the University of Michigan in 1964. At the time of his death in 1996, he was the Davis Brack Professor of Behavioral Sciences in the Department of Psychology at Stanford University. Previously he held professorships at the Hebrew University of Jerusalem and Harvard University. A fellow at the Center for Advanced Study in 1970, he was elected to the American Academy of Arts and Sciences in 1980 and the National Academy of Science in 1985. He also won (with Kahneman) the American Psychological Association's award for distinguished scientific contribution in 1982, and MacArthur and Guggenheim fellowships in 1984. He was awarded honorary doctorates by the University of Chicago, Yale University, the University of Goteborg in Sweden, and the State University of New York at Buffalo.

JANET YELLEN is currently the Eugene E. and Catherine M. Trefethen Professor of Business Administration at the Haas School of Business and Professor in the Department of Economics at the University of California, Berkeley. She served as the chair of the President's Council of Economic Advisors in the Clinton administration, and was a member of the Board of Governors of the Federal Reserve System from 1994 to 1997.

P R E F A C E

THIS BOOK was conceived several years ago when the editors, along with Drazen Prelec and Dick Thaler, spent a year as a working group at the Center for Advanced Study in Behavioral Sciences (CASES). When we weren't playing volleyball or hiking, we spent a lot of time taking stock of our field, making lists of what the main contributions were, and idly speculating about the future. We also contemplated various group projects, such as coediting a *Handbook of Behavioral Economics*. But none of us wanted to commit the time and energy it would take to ride herd on a group of authors who regard procrastination as such a regular feature of human behavior that they would be unembarrassed to procrastinate themselves. So the idea of a book of readings emerged, and eventually evolved into a collection of recent, important papers in the field.

The title of this collection deliberately bears the word "Advances" because we omitted many classic articles (which, by the way, any serious student of behavioral economics should read; our introductory chapter is partly designed to be an annotated guide to these influential classics). Including all of the deserving classic articles and newer contributions in one volume just stretched coverage of either type of article too thin. Fortunately, the early classics are available in many other places, including Kahneman, Slovic, and Tversky *Judgment under Uncertainty: Heuristics and Biases* (1982) on judgment; Kahneman and Tversky *Choices, Values and Frames* (2001) on choice; Elster and Loewenstein, *Choice over Time* (1992) on intertemporal choice; and Thaler's essential *The Winner's Curse* (1992). More recent compilations include Gilovich, Kahneman, and Miller, *Heuristics of Judgment: Extensions and Applications* (2002) on judgment; and Loewenstein, Read, and Baumeister, *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice* (2003) on the latest thinking about intertemporal choice.

The fact that we had to make a hard choice, and leave so many worthy papers out of the volume—not only classics, but also current works—is a testament to the progress of the field. Twenty years ago, behavioral economics did not exist as a field. There were scattered works by authors such as Duesenberry, Galbraith, Katona, Leibenstein, and Scitovsky, which received attention, but the general attitude of the field toward psychology was one of hostility and skepticism. Many economists simply didn't think it was necessary to try to model psychological limits (since errors would be extinguished by market, advice, evolution, etc.), or that it was even possible to do so parsimoniously. The older two of us experienced this hostility first-hand, from faculty members during graduate school, and later even more extremely when we attempted to publish. In fact, until about 1990, it was not uncommon to get a paper returned from a journal (usually after a delay of about a year) with a three sentence referee report saying "this isn't economics." Fortunately, hostility switched to curiosity and acceptance rather rapidly and completely in the past few years.

How did we get here from there? A big part of the credit should go to the people to whom this book is dedicated. Kahneman and Tversky provided the raw materials for much of behavioral economics—a new line of psychology, called behavioral decision research, that draws explicit contrasts between descriptively realistic accounts of judgment and choice and the assumptions and predictions of economics. Richard Thaler was the first economist to recognize the potential applications of this research to economics. His 1980 article "Toward a theory of consumer choice," published in the first issue of the remarkably open-minded (for its time) *Journal of Economic Behavior and Organization*, is considered by many to be the first genuine article in modern behavioral economics. (Thaler's 1999 article, which updates the earlier one and extends it, is included here in *Advances*.) Thaler's "anomalies" column published in the *Journal of Economic Perspectives* was another critical element in getting people to pay attention to behavioral economics. The anomalies column helped to shift many economists from the attitude "if it works don't try to fix it" to "it's broken; how can we fix it?"

Needless to say, numerous other scholars played important roles, including the psychologists Ward Edwards, Hillel Einhorn, Baruch Fischhoff, Robin Hogarth, Ken Hammond, Sarah Lichtenstein, and Paul Slovic. Herb Simon—the only psychologist before Kahneman to win the Nobel prize in economics—coined the terms "bounded rationality" and "procedural rationality" and urged economists to model the implications of bounds and procedures.

Behavioral economics also flourished because it was encouraged and done early on by economists who were better-known for other kinds of work, including George Akerlof, Ken Arrow, Peter Diamond, Bob Shiller, Lawrence Summers, Sidney Winter, and Richard Zeckhauser. (Our apologies for omitting many other important contributors in these lists. Can we plead guilty to "availability" bias?)

All these scientists played important roles in the advancement of behavioral economics. Our dedication includes one other person who played an unusual and vital role—Eric Wanner, the president of the Russell Sage Foundation. Wanner was first exposed to behavioral economics in the mid-1980s as a program officer at the Sloan Foundation. Sloan sponsored a small conference on psychology and economics that was attended by two of us (Camerer and Loewenstein) Kahneman, Tversky, Thaler, and others. While Sloan did not bet heavily on the emerging field, Wanner did make a big bet after taking the job of president of the Russell Sage Foundation (RSF).

RSF's official charge is to fund social science research to help the poor. Wanner, an accomplished cognitive psychologist early in his career, felt that rational-choice economics provided a limited scientific language in which to talk about sources of poverty and about policy solutions. He saw in behavioral economics the chance for a small foundation to have a big impact in social science and to broaden the language of economics to say more about poverty. He funded research in behavioral economics and invited many behavioral economists to the foundation as fellows in residence, including two of us (Camerer and Loewenstein).

A brilliant RSF investment was a series of biannual "summer camps," started in 1994 to teach behavioral economics to advanced graduate students in economics and other social sciences. Like other summer camps in economics, these have been hugely effective in conveying a body of knowledge that campers could not get in

Ph.D. courses at their home schools, until recently. The rosters of guest speakers and camper alumni are both impressive indeed. The camps have also sharpened our own thinking, and created a social network of students from around the world.

The most recent program of RSF's support for behavioral economics has been the copublication, with Princeton University Press, of a Behavioral Economics Roundtable Series. This book is the second of many planned volumes in that series. The field is progressing so rapidly that an advanced *Advances* is not far away.

A C K N O W L E D G M E N T S

THE FOLLOWING publishers kindly granted permission to reprint the material that appears in the book.

Chapter 2: Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy.* © 1990 by the University of Chicago Press.

Chapter 3: Richard H. Thaler. "Mental Accounting Matters." *Journal of Behavioral Decision Making*. © 1999 John Wiley and Sons Limited.

Chapter 4: Chris Starmer. "Developments in Nonexpected-Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk." *Journal of Economic Literature*. © 2000 American Economic Association.

Chapter 5: Colin F. Camerer. Prospect Theory in the Wild. In *Choices, Values and Frames*, edited by Daniel Kahneman and Amos Tversky. © 2000 Russell Sage Foundation and Cambridge University Press.

Chapter 6: Shane Frederick, George Loewenstein, and Ted O'Donoghue. "Time Discounting: A Critical Review." *Journal of Economic Literature*. © 2002 American Economic Association.

Chapter 7: Ted O'Donoghue and Matthew Rabin. "Doing it Now or Later." *The American Economic Review*. © 1999 American Economic Association.

Chapter 8: Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review*. © 1986 American Economic Association.

Chapter 9: Ernst Fehr and Klaus M. Schmidt. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics*. © 1999 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

Chapter 10: Matthew Rabin. "Incorporating Fairness into Game theory and Economics," *American Economic Review*. © 1993 American Economic Association.

Chapter 11: Linda Babcock and George Loewenstein. "Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives*. © 1997 American Economic Association.

Chapter 12: Vincent P. Crawford. "Theory and Experiment in the Analysis of Strategic Interaction." In *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress*, Vol 1, edited by D. Kreps and K. Wallis. © 1997 Cambridge University Press.

Chapter 14: Hersh M. Shefrin and Richard H. Thaler. "Mental Accounting, Saving, and Self-Control." In *Choice Over Time*, edited by George Loewenstein and Jon Elster. © 1992 Russell Sage Foundation.

Chapter 15: David Laibson. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*. © 1997 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

Chapter 16: George A. Akerlof and Janet L. Yellen. "The Fair Wage-Effort Hypothesis and Unemployment." *The Quarterly Journal .of Economics*. © 1990 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

Chapter 17: Eldar Shafir, Peter Diamond, and Amos Tversky. “Money Illusion.” *Quarterly Journal of Economics*. © 1997 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

Chapter 18: Ernst Fehr and Simon Gächter. “Fairness and Retaliation: The Economics of Reciprocity.” *Journal of Economic Perspectives*. © 2001 American Economic Association.

Chapter 19: Colin Camerer, Linda Babcock, George Loewenstein, and Richard H. Thaler. “Labor Supply of New York City Cab Drivers: One Day at a Time.” *Quarterly Journal of Economics*. © 1997 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

Chapter 20: Robert H. Frank and Robert M. Hutchens. “Wages, Seniority, and the Demand for Rising Consumption Profiles.” *Journal of Economic Behavior and Organization*. © 1993 Elsevier Science.

Chapter 22: Shlomo Benartzi and Richard H. Thaler. “Myopic Loss-Aversion and the Equity Premium Puzzle.” *Quarterly Journal of Economics*. © 1995 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

Chapter 23: Terrance Odean. “Do Investors Trade Too Much?” *American Economic Review*. © 1999 American Economic Association.

Chapter 24: David Genesove and Christopher Mayer. “Loss-Aversion and Seller Behavior: Evidence from the Housing Market. *Quarterly Journal of Economics*. © 2001 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

Chapter 25: Itzhak Gilboa and David Schmeidler. “Case-Based Decision Theory.” *Quarterly Journal of Economics*. © 1995 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

Chapter 26: George Loewenstein. “Out of Control: Visceral Influences on Decision Making.” In *Organizational Behavior and Human Decision Processes*, Volume 65. © 1996 Elsevier Science (USA).

PART I

Introduction

CHAPTER 1

Behavioral Economics: Past, Present, Future

COLIN F. CAMERER AND GEORGE LOEWENSTEIN

Behavioral economics increases the explanatory power of economics by providing it with more realistic psychological foundations. This book consists of representative recent articles in behavioral economics.[1] Chapter 1 is intended to provide an introduction to the approach and methods of behavioral economics, and to some of its major findings, applications, and promising new directions. It also seeks to fill some unavoidable gaps in the chapters' coverage of topics.

What Behavioral Economics Tries to Do

At the core of behavioral economics is the conviction that increasing the realism of the psychological underpinnings of economic analysis will improve the field of economics *on its own terms*—generating theoretical insights, making better predictions of field phenomena, and suggesting better policy. This conviction does not imply a wholesale rejection of the neoclassical approach to economics based on utility maximization, equilibrium, and efficiency. The neoclassical approach is useful because it provides economists with a theoretical framework that can be applied to almost any form of economic (and even noneconomic) behavior, and it makes refutable predictions. Many of these predictions are tested in the chapters of this book, and rejections of those predictions suggest new theories.

Most of the papers modify one or two assumptions in standard theory in the direction of greater psychological realism. Often these departures are not radical at all because they relax simplifying assumptions that are not central to the economic approach. For example, there is nothing in core neoclassical theory that specifies that people should not care about fairness, that they should weight risky outcomes in a linear fashion, or that they must discount the future exponentially at a constant rate.[2] Other assumptions simply acknowledge human limits on computational

We thank Steve Burks, Richard Thaler, and especially Matthew Rabin (who collaborated during most of the process) for the helpful comments.

[1] Since it is a book of advances, many of the seminal articles that influenced those collected here are not included, but are noted below and are widely reprinted elsewhere.

[2] While the chapters in this book largely adhere to the basic neoclassical framework, there is nothing inherent in behavioral economics that *requires* one to embrace the neoclassical economic model. Indeed, we consider it likely that alternative paradigms will eventually be proposed that have greater explanatory power. Recent developments in psychology, such as connectionist models that capture

power, willpower, and self-interest. These assumptions can be considered "procedurally rational" (Herbert Simon's term) because they posit functional heuristics for solving problems that are often so complex that they cannot be solved exactly by even modern computer algorithms.

Evaluating Behavioral Economics

Stigler (1965) says economic theories should be judged by three criteria: congruence with reality, generality, and tractability. Theories in behavioral economics should be judged this way too. We share the modernist view that the ultimate test of a theory is the accuracy with which it identifies the actual causes of behavior; making accurate predictions is a big clue that a theory has pinned down the right causes, but more *realistic* assumptions are surely helpful too.[3]

Theories in behavioral economics also strive for *generality*—e.g., by adding only one or two parameters to standard models. Particular parameter values then often reduce the behavioral model to the standard one, and the behavioral model can be pitted against the standard model by estimating parameter values. Once parameter values are pinned down, the behavioral model can be applied just as widely as the standard one.

Adding behavioral assumptions often *does* make the models less tractable. However, many of the papers represented in this volume show that it can be done. Moreover, despite the fact that they often add parameters to standard models, behavioral models, in some cases, can be even more *precise* than traditional ones that assume more rationality, when there is dynamics and strategic interaction. Thus, Lucas (1986) noted that rational expectations allow for multiple inflationary and asset price paths in dynamic models, while adaptive expectations pin down one path. The same is true in game theory: Models based on cognitive algorithms (Camerer, Ho, and Chong 2003) often generate precise predictions in those games where the mutual consistency requirement of Nash permits multiple equilibria.

The realism, generality, and tractability of behavioral economics can be illustrated with the example of loss-aversion. Loss-aversion is the disparity between the strong aversion to losses relative to a reference point and the weaker desire for gains of equivalent magnitude. Loss aversion is more *realistic* than the standard continuous, concave, utility function over wealth, as demonstrated by hundreds of experiments. Loss aversion has proved useful in identifying where predictions of standard theories will go wrong: Loss-aversion can help account for the equity premium puzzle in finance and asymmetry in price elasticities. (We provide more examples further on.) Loss aversion can also be parameterized in a general way, as the ratio of the marginal disutility of a loss relative to the marginal utility of a

---

some of the essential features of neural functioning, bear little resemblance to models based on utility maximization, yet are reaching the point where they are able to predict many judgmental and behavioral phenomena.

[3] Contrary to the positivistic view, however, we believe that predictions of feelings (e.g., of subjective well-being) should also be an important goal.

gain at the reference point (i.e., the ratio of the derivatives at zero); the standard model is the special case in which this "loss-aversion coefficient" is 1. As the foregoing suggests, loss-aversion has proved *tractable*—although not always simple—in several recent applications (Barberis, Huang, and Santos 2001).

The Historical Context of Behavioral Economics

Most of the ideas in behavioral economics are not new; indeed, they return to the roots of neoclassical economics after a century-long detour. When economics first became identified as a distinct field of study, psychology did not exist as a discipline. Many economists moonlighted as the psychologists of their times. Adam Smith, who is best known for the concept of the "invisible hand" and *The Wealth of Nations*, wrote a less well-known book, *The Theory of Moral Sentiments*, which laid out psychological principles of individual behavior that are arguably as profound as his economic observations. The book is bursting with insights about human psychology, many of which presage current developments in behavioral economics. For example, Adam Smith commented (1759 / 1892, 311) that "we suffer more . . . when we fall from a better to a worse situation, than we ever enjoy when we rise from a worse to a better." Loss aversion! Jeremy Bentham, whose utility concept formed the foundation of neoclassical economics, wrote extensively about the psychological underpinnings of utility, and some of his insights into the determinants of utility are only now starting to be appreciated (Loewenstein 1999). Francis Edgeworth's *Theory of Mathematical Psychics* introduced his famous "box" diagram showing two-person bargaining outcomes and included a simple model of social utility, in which one person's utility was affected by another person's payoff, which is a springboard for modern theories (see chapters 9 and 10 of this volume—*Advances in Behavioral Economics*—for two examples).

The rejection of academic psychology by economists, perhaps somewhat paradoxically, began with the neoclassical revolution, which constructed an account of economic behavior built up from assumptions about the nature—that is, the *psychology*—of homo economicus. At the turn of the twentieth century, economists hoped that their discipline could be like a natural science. Psychology was just emerging at that time and was not very scientific. The economists thought it provided too unsteady a foundation for economics. Their distaste for the psychology of their period, as well as their dissatisfaction with the hedonistic assumptions of Benthamite utility, led to a movement to expunge the psychology from economics.[4]

[4] The economists of the time had less disagreement with psychology than they realized. Prominent psychologists of the time were united with the economists in rejecting hedonism as the basis of behavior. William James, for example, wrote that "psychologic hedonists obey a curiously narrow teleological superstition, for they assume without foundation that behavior always aims at the *goal* of maximum pleasure and minimum pain; but behavior is often impulsive, not goal-oriented," while William McDougall stated in 1908 that "it would be a libel, not altogether devoid of truth, to say that classical political economy was a tissue of false conclusions drawn from false psychological assumptions." Both quotes from Lewin (1996).

The expunging of psychology from economics happened slowly. In the early part of the twentieth century, the writings of economists such as Irving Fisher and Vilfredo Pareto still included rich speculations about how people feel and think about economic choices. Later, John Maynard Keynes appealed frequently to psychological insights, but by the middle of the century discussions of psychology had largely disappeared.

Throughout the second half of the century, many criticisms of the positivistic perspective took place in both economics and psychology. In economics, researchers like George Katona, Harvey Leibenstein, Tibor Scitovsky, and Herbert Simon wrote books and articles suggesting the importance of psychological measures and bounds on rationality. These commentators attracted attention but did not alter the fundamental direction of economics.

Many coincidental developments led to the emergence of behavioral economics as represented in this book. One development was the rapid acceptance by economists of the expected utility and discounted utility models as normative and descriptive models of decision making under uncertainty and intertemporal choice, respectively. Whereas the assumptions and implications of generic utility analysis are rather flexible, and hence tricky to refute, the expected utility and discounted utility models have numerous precise and testable implications. As a result, they provided some of the first "hard targets" for critics of the standard theory. Seminal papers by Allais (1953), Ellsberg (1961), and Markowitz (1952) pointed out anomalous implications of expected and subjective expected utility. Strotz (1955) questioned exponential discounting. Later scientists demonstrated similar anomalies using compelling experiments that were easy to replicate (Kahneman and Tversky 1979, on expected utility; Thaler 1981, and Loewenstein and Prelec 1992, on discounted utility).

As economists began to accept anomalies as counterexamples that could not be permanently ignored, developments in psychology identified promising directions for new theory. Beginning around 1960, cognitive psychology became dominated by the metaphor of the brain as an information-processing device, which replaced the behaviorist conception of the brain as a stimulus-response machine. The information-processing metaphor permitted a fresh study of neglected topics like memory, problem solving and decision making. These new topics were more obviously relevant to the neoclassical conception of utility maximization than behaviorism had appeared to be. Psychologists such as Ward Edwards, Duncan Luce, Amos Tversky, and Daniel Kahneman began to use economic models as a benchmark against which to contrast their psychological models. Perhaps the two most influential contributions were published by Tversky and Kahneman. Their 1974 *Science* article argued that heuristic short-cuts created probability judgments that deviated from statistical principles. Their 1979 paper "Prospect theory: Decision making under risk" documented violations of expected utility and proposed an axiomatic theory, grounded in psychophysical principles, to explain the violations. The latter was published in the technical journal *Econometrica* and is one of the most widely cited papers ever published in that journal.

A later milestone was the 1986 conference at the University of Chicago, at which an extraordinary range of social scientists presented papers (see Hogarth and Reder 1987). Ten years later, in 1997, a special issue of the *Quarterly Journal of Economics* was devoted to behavioral economics (three of those papers are reprinted in this volume).

Early papers established a recipe that many lines of research in behavioral economics have followed. First, identify normative assumptions or models that are ubiquitously used by economists, such as Bayesian updating, expected utility, and discounted utility. Second, identify anomalies—i.e., demonstrate clear violations of the assumption or model, and painstakingly rule out alternative explanations, such as subjects' confusion or transactions costs. And third, use the anomalies as inspiration to create alternative theories that generalize existing models. A fourth step is to construct economic models of behavior using the behavioral assumptions from the third step, derive fresh implications, and test them. This final step has only been taken more recently but is well represented in this volume of advances.

The Methods of Behavioral Economics

The methods used in behavioral economics are the same as those in other areas of economics. At its inception, behavioral economics relied heavily on evidence generated by experiments. More recently, however, behavioral economists have moved beyond experimentation and embraced the full range of methods employed by economists. Most prominently, a number of recent contributions to behavioral economics, including several included in this book (chapters 21, 25, and 26, and studies discussed in chapters 7 and 11) rely on field data. Other recent papers utilize methods such as field experiments (Gneezy and Rustichini, in this volume) computer simulation (Angeletos et al. 2001), and even brain scans (McCabe et al. 2001).

Experiments played a large role in the initial phase of behavioral economics because experimental control is exceptionally helpful for distinguishing behavioral explanations from standard ones. For example, players in highly anonymous one-shot take-it-or-leave-it "ultimatum" bargaining experiments frequently reject substantial monetary offers, ending the game with nothing (see Camerer and Thaler 1995). Offers of 20% or less of a sum are rejected about half the time, even when the amount being divided is several weeks' wages or $400 (U.S.) (Camerer 2003). Suppose we observed this phenomenon in the field, in the form of failures of legal cases to settle before trial, costly divorce proceedings, and labor strikes. It would be difficult to tell whether rejection of offers was the result of reputation-building in repeated games, agency problems (between clients and lawyers), confusion, or an expression of distaste for being treated unfairly. In ultimatum game experiments, the first three of these explanations are ruled out because the experiments are played once anonymously, have no agents, and are simple enough to rule out confusion. Thus, the experimental data clearly establishes that subjects are expressing concern for fairness. Other experiments have been useful for testing whether judgment errors that individuals commonly make in psychology experiments also affect prices and quantities in markets. The lab is especially useful for these studies because individual and market-level data can be observed simultaneously (Camerer 1987; Ganguly, Kagel, and Moser 2000).

Although behavioral economists initially relied extensively on experimental data, we see behavioral economics as a very different enterprise from experimental economics (see Loewenstein 1999). As noted, behavioral economists are methodological eclectics. They define themselves not on the basis of the research methods that they employ but rather on their application of psychological insights to economics. Experimental economists, on the other hand, define themselves on the basis of their endorsement and use of experimentation as a research tool. Consistent with this orientation, experimental economists have made a major investment in developing novel experimental methods that are suitable for addressing economic issues and have achieved a virtual consensus among themselves on a number of important methodological issues.

This consensus includes features that we find appealing and worthy of emulation (see Hertwig and Ortmann, 2001). For example, experimental economists often make instructions and software available for precise replication, and raw data are typically archived or generously shared for reanalysis. Experimental economists insist on paying performance-based incentives, which reduces response noise (but does not typically improve rationality; see Camerer and Hogarth 1999), and also have a prohibition against deceiving subjects.

However, experimental economists have also developed rules that many behavioral economists are likely to find excessively restrictive. For example, experimental economists rarely collect data like demographics, self-reports, response times, and other cognitive measures that behavioral economists have found useful. Descriptions of the experimental environment are usually abstract rather than evocative of a particular context in the outside world because economic theory rarely makes a prediction about how contextual labels would matter, and experimenters are concerned about losing control over incentives if choosing strategies with certain labels is appealing because of the labels themselves. Psychological research shows that the effect of context on decision making can be powerful (see Goldstein and Weber 1995; Loewenstein 2001) and some recent experimental economics studies have explored context effects too (Cooper et al. 1999; Hoffman et al. 1994). Given that context is likely to matter, the question is whether to treat it as a nuisance variable or an interesting treatment variable. It is worth debating further whether or not it is useful to help subjects see a connection between the experiment and the naturally occurring situations the experiment is designed to model, by using contextual cues.

Economics experiments also typically use "stationary replication"—in which the same task is repeated over and over, with fresh endowments in each period. Data from the last few periods of the experiment are typically used to draw conclusions about equilibrium behavior outside the lab. While we believe that examining behavior after it has converged is of great interest, it is also obvious that many important aspects of economic life are like the *first* few periods of an experiment rather than the *last.* If we think of marriage, educational decisions, saving for retirement, or the purchase of large durables like houses, sailboats, and cars, which happen just a few times in a person's life, a focus exclusively on "post-convergence" behavior is clearly not warranted.[5]

All said, the focus on psychological realism and economic applicability of research promoted by the behavioral-economics perspective suggests the immense usefulness of both empirical research outside the lab and of a broader range of approaches to laboratory research.

Basic Concepts and Research Findings

The field of behavioral decision research, on which behavioral economics has drawn more than any other subfield of psychology, typically classifies research into two categories: judgment and choice. Judgment research deals with the processes that people use to estimate probabilities. Choice deals with the processes people use to select among actions, taking account of any relevant judgments that they may have made. In this section, we provide a background on these two general topics to put the contributions of specific chapters into a broader context.

Probability Judgment

Judging the likelihood of events is central to economic life. Will you lose your job in a downturn? Will you be able to find another house you like as much as the one you must bid for right away? Will the Fed raise interest rates? Will an AOL-TimeWarner merger increase profits? Will it rain during your vacation to London? These questions are answered by some process of judging likelihood.

The standard principles used in economics to model probability judgment in economics are concepts of statistical sampling, and Bayes's rule for updating probabilities in the face of new evidence. Bayes's rule is unlikely to be correct descriptively because it has several features that are cognitively unrealistic. First, Bayesian updating requires a prior.[6] Second, Bayesian updating requires a separation between previously judged probabilities and evaluations of new evidence. But many cognitive mechanisms use previous information to filter or interpret what is observed, violating this separability. For example, in perception experiments, subjects who expect to see an object in a familiar place—such as a fire hydrant on a sidewalk—perceive that object more accurately than subjects who see the same object in an unexpected place—such as on a coffeeshop counter. Third, subjective expected utility assumes separability between probability judgments of states and utilities that result from those states. Wishful thinking and

[5] We call the standard approach "Groundhog Day" replication, after the Bill Murray movie in which the hero finds himself reliving exactly the same day over and over. Murray's character is depressed until he realizes that he has the ideal opportunity to learn by trial-and-error, in a stationary environment, and uses the opportunity to learn how to woo his love interest.

[6] Because it does not specify where the prior comes from, however, it leaves room for psychological theory on the front end of the judgment process.

other self-serving motivations violate this separation (see Babcock and Loewenstein 1997 and in this volume). Fourth, the Bayesian updating predicts no effects of the order of arrival of information. But, order effects are common in memory due to the strength of recent information in working memory (recency effects) and of increased "rehearsal" of older memories (primacy effects). These order effects mean that how information is sequenced distorts probability judgment (see Hogarth and Einhorn 1992).

Cognitive psychologists have proposed heuristic mechanisms that will lead to judgments which sometimes violate either sampling principles or Bayes's rule (see Kahneman and Frederick 2002). For example, people may judge the probabilities of future events based on how easy those events are to imagine or to retrieve from memory. This "availability heuristic" contributes to many specific further biases. One is "hindsight bias": Because events that actually occurred are easier to imagine than counterfactual events that did not, people often overestimate the probability they previously attached to events that later happened. This bias leads to "second guessing" or Monday-morning quarterbacking and may be partly responsible for lawsuits against stockbrokers who lost money for their clients. (The clients think that the brokers "should have known.") A more general bias is called the "curse of knowledge"—people who know a lot find it hard to imagine how little others know. The development psychologist Jean Piaget suggested that the difficulty of teaching is caused by this curse. (For example, why is it so hard to explain something "obvious" like consumer indifference curves or Nash equilibrium to your undergraduate students?[7]) Anybody who has tried to learn from a computer manual has seen the curse of knowledge in action.

Another heuristic for making probability judgments is called "representativeness": People judge conditional probabilities like P (hypothesis / data) or P (example / class) by how well the data represents the hypothesis or the example represents the class. Like most heuristics, representativeness is an economical shortcut that delivers reasonable judgments with minimal cognitive effort in many cases, but sometimes goofs badly and is undisciplined by normative principles. Prototypical exemplars of a class may be judged to be more likely than they truly are (unless the prototype's extremity is part of the prototype). For example, in judging whether a certain student described in a profile is, say, a psychology major or a computer science major, people instinctively dwell on how well the profile matches the psychology or computer science major stereotype. Many studies show how this sort of feature-matching can lead people to underweigh the "base rate"—in this example, the overall frequency of the two majors.[8]

[7] Here is an example from the business world: When its software engineers refused to believe that everyday folks were having trouble learning to use their opaque, buggy software, Microsoft installed a test room with a one-way mirror so that the engineers could see people struggling before their very eyes (Heath, Larrick, and Klayman 1998).

[8] However, this "base-rate fallacy" is being thoughtfully reexamined (Koehler 1996). The fact that base rates are more clearly included when subjects are asked what fraction of 100 hypothetical cases fit the profile is an important clue about how the heuristic operates and its limits (Gigerenzer, Hell, and Blank 1988; Tversky and Kahneman 1983).

Another by-product of representativeness is the "law of small numbers." Small samples are thought to represent the properties of the statistical process that generated them (as if the law of large numbers, which guarantees that a large sample of independent draws *does* represent the process, is in a hurry to work). If a baseball player gets hits 30% of his times at bat, but is 0 for 4 so far in a particular game, then he is "due" for a hit in his next time at bat in this game, so that this game's hitting profile will more closely represent his overall ability. The so-called "gambler's fallacy," whereby people expect a tail after a coin landed heads three times in a row, is one manifestation of the law of small numbers. The flip side of the same misjudgment (so to speak) is surprise at the long streaks that result if the time series is random, which can lead people to conclude that the coin must be unfair when it isn't. Field and experimental studies with basketball shooting and betting on games show that people, including bettors, believe that there is positive autocorrelation—that players experience the "hot hand"—when there is no empirical evidence that such an effect exists (see Camerer 1989a; Gilovich, Vallone, and Tversky 1985).

Many studies explore these heuristics and replicate their "biases" in applied domains (such as judgments of accounting auditors, consumers buying products, and students in classroom negotiations). It is important to note that a "heuristic" is both a good thing and a bad thing. A good heuristic provides fast, close to optimal, answers when time or cognitive capabilities are limited, but it also violates logical principles and leads to errors in some situations. A lively debate has emerged over whether heuristics should be called irrational if they were well-adapted to domains of everyday judgment ("ecologically rational"). In their early work, Kahneman, Tversky, and others viewed cognitive biases as the judgmental kin of speech errors ("I cossed the toin"), forgetting, and optical illusions: These are systematic errors that, even if rare, are useful for illuminating how cognitive mechanisms work. But these errors do not imply that the mechanisms fail frequently or are not well adapted for everyday use. But as Kahneman and Tversky (1982, p. 494) wrote, "Although errors of judgment are but a method by which some cognitive processes are studied, the method has become a significant part of the message." The shift in emphasis from the heuristics to the biases that they sometimes create happened gradually as research moved to applied areas; the revisionist view that heuristics may be near-optimal is largely a critique (a reasonable one) of the later applied research.

Progress in modeling and applying behavioral models of judgment has lagged behind other areas, such as loss aversion and hyperbolic time discounting. A promising recent modeling approach is "quasi-Bayesian"—viz., assume that people misspecify a set of hypotheses, or encode new evidence incorrectly, but otherwise use Bayes's rule. For example, Rabin and Schrag (1999) model "confirmation bias" by assuming that people who believe hypothesis A is more likely than B will never encode pro-A evidence mistakenly, but will sometimes encode pro-B evidence as being supportive of A.[9] Rabin (2002) models the "law of small numbers"

[9] This encoding asymmetry is related to "feature-positive" effects and perceptual encoding biases that are well documented in research on perception. After buying a Volvo, you will suddenly "see" more Volvos on the road, due purely to heightened familiarity.

in a quasi-Bayesian fashion by assuming that people mistakenly think that a process generates draws from a hypothetical "urn" *without replacement*, although draws are actually independent (i.e., made *with* replacement). He shows some surprising implications of this misjudgment. For example, investors will think that there is wide variation in skill of, say, mutual-fund managers, even if there is no variation at all. A manager who does well several years in a row is a surprise if performance is mistakenly thought to be mean-reverting due to "nonreplacement," so quasi-Bayesians conclude that the manager must be *really* good.

Barberis, Shleifer, and Vishny (1998) adopt such a quasi-Bayesian approach to explain why the stock market underreacts to information in the short-term and overreacts in the long-term. In their model, earnings follow a random walk but investors believe, mistakenly, that earnings have positive momentum in some regimes and regress toward the mean in others. After one or two periods of good earnings, the market can't be confident that momentum exists and hence expects mean-reversion; but since earnings are really a random walk, the market is too pessimistic and is underreacting to good earnings news. After a long string of good earnings, however, the market believes momentum is building. Since it isn't, the market is too optimistic and overreacts.

While other approaches that discover ways of formalizing some of the findings of cognitive psychology are possible, our guess is that the quasi-Bayesian view will quickly become the standard way of translating the cognitive psychology of judgment into a tractable alternative to Bayes's rule. The models mentioned in the previous two paragraphs are parameterized in such a way that the Bayesian model is embedded as a special case, which allows theoretical insight and empirical tests about how well the Bayesian restriction fits.

Preferences: Revealed, Constructed, Discovered, or Learned?

Standard preference theory incorporates a number of strong and testable assumptions. For example, it assumes that preferences are "reference independent"—i.e., they are not affected by the individual's transient asset position. It also assumes that preferences are invariant with respect to superficial variations in the way that options are described, and that elicited preferences do not depend on the precise way in which preferences are measured as long as the method used is "incentive compatible"—i.e., provides incentives for people to reveal their "true" preferences. All of these assumptions have been violated in significant ways (see Slovic 1995).

For example, numerous "framing effects" show that the way that choices are presented to an individual often determine the preferences that are "revealed." The classic example of a framing effect is the "Asian disease" problem in which people are informed about a disease that threatens 600 citizens and asked to choose between two undesirable options (Tversky and Kahneman 1981). In the "positive frame," people are given a choice between (A) saving 200 lives for sure, or (B) a one-third chance of saving all 600 with a two-third chance of saving no one. In the "negative frame," people are offered a choice between (C) 400 people dying for sure, or (D) a two-third chance of 600 dying and a one-third chance of no one dying. Despite the fact that A and C, and B and D, are equivalent in terms of lives lost or at risk, most people choose A over B but D over C.

Another phenomenon that violates standard theory is called an "anchoring effect." The classic demonstration of an anchoring effect (Tversky and Kahneman 1974 and in this volume) was identified in the context of judgment rather than choice. Subjects were shown the spin of a wheel of fortune that could range between 0 and 100 and were asked to guess whether the number of African nations in the United Nations was greater than or less than this number. They were then asked to guess the true value. Although the wheel of fortune was obviously random, subjects' guesses were strongly influenced by the spin of the wheel. As Kahneman and Tversky interpreted it, subjects seemed to "anchor" on the number spun on the wheel and then adjusted for whatever else they thought or knew, but adjusted insufficiently. Of interest in this context is that anchoring effects have also been demonstrated for choices as opposed to judgments. In one study, subjects were asked whether their certainty equivalent for a gamble was greater than or less than a number chosen at random and then were asked to specify their actual certainty equivalent for the gamble (Johnson and Schkade 1989). Again, the stated values were correlated significantly with the random value.

In a recent study of anchoring, Ariely, Loewenstein, and Prelec (2003) sold valuable consumer products (a $100 wireless keyboard, a fancy computer mouse, bottles of wine, and a luxurious box of chocolate) to postgraduate (MBA) business students. The students were presented with a product and asked whether they would buy it for a price equal to the last two digits of their own social security number (a roughly random identification number required to obtain work in the United States) converted into a dollar figure—e.g., if the last digits were 79, the hypothetical price was $79. After giving a yes / no response to the question "Would you pay $79?" subjects were asked to state the most they would pay (using a procedure that gives people an incentive to say what they really would pay). Although subjects were reminded that the social security number is essentially random, those with high numbers were willing to pay more for the products. For example, subjects with numbers in the bottom half of the distribution priced a bottle of wine— a 1998 Côtes du Rhône Jaboulet Parallel '45—at $11.62, while those with numbers in the top half priced the same bottle at $19.95.

Many studies have also shown that the *method* used to elicit preferences can have dramatic consequences, sometimes producing "preference reversals"—situations in which A is preferred to B under one method of elicitation, but A is judged as inferior to B under a different elicitation method (Grether and Plott 1979). The best-known example contrasts how people choose between two bets versus what they separately state as their selling prices for the bets. If bet A offers a high probability of a small payoff and bet B offers a small probability of a high payoff, the standard finding is that people choose the more conservative A bet over bet B when asked to choose, but are willing to pay more for the riskier bet B when asked to price them separately. Another form of preference reversal occurs between joint and separate evaluations of pairs of goods (Hsee et al. 1999; see

Hsee and LeClerc [1998] for an application to marketing). People will often price or otherwise evaluate an item A higher than another item B when the two are evaluated independently, but evaluate B more highly than A when the two items are compared and priced at the same time.

"Context effects" refer to ways in which preferences between options depend on what other options are in the set (contrary to "independence of irrelevant alternatives" assumptions). For example, people are generally attracted to options that dominate other options (Huber, Payne, and Puto 1982). They are also drawn disproportionately to "compromise" alternatives with attribute values that lie between those of other alternatives (Simonson and Tversky 1992).

All of the above findings suggest that preferences are not the predefined sets of indifference curves represented in microeconomics textbooks. They are often ill-defined, highly malleable, and dependent on the context in which they are elicited. Nevertheless, when required to make an economic decision—to choose a brand of toothpaste, a car, a job, or how to invest—people do make *some* kind of decision. Behavioral economists refer to the process by which people make choices with ill-defined preferences as "constructing preferences" (Payne, Bettman, and Johnson 1992; Slovic 1995).

A theme emerging in recent research is that, although people often reveal inconsistent or arbitrary preferences, they typically obey normative principles of economic theory when it is transparent how to do so. Ariely, Loewenstein, and Prelec (2003) refer to this pattern as "coherent arbitrariness" and illustrate the phenomenon with a series of studies in which the amount of money subjects must be paid to listen to an annoying sound is sensitive to an arbitrary anchor, but they also must be paid much more to listen to the tone for a longer period of time. Thus, while expressed valuations for one unit of a good are sensitive to an anchor that is clearly arbitrary, subjects also obey the normative principle of adjusting those valuations to the quantity—in this case, the duration—of the annoying sound.[10]

Most evidence that preferences are constructed comes from demonstrations that a feature that should not matter actually does. The way in which gambles are "framed" as gains and losses from a reference outcome, in which the composition of a choice is set, and whether people choose among objects or value them separately, have all been shown to make a difference in expressed preference. But admittedly, a list of a theory's failings is not an alternative theory. So far, a parsimonious alternative theory has not emerged to deal with all of these challenges to utility maximization.[11]

[10] A joke makes this point nicely. An accountant flying across the country nudges the person in the next seat. "See those mountains down there?" the accountant asks. "They're a million and four years old." Intrigued, the neighbor asks how the accountant can be so sure of the precise age of the mountains. The accountant replied, "Well, four years ago I flew across these mountains and a geologist I sat next to said they were a million years old. So now they're a million and four."

[11] Some specialized models have been proposed to explain particular phenomena, such as Hsee, Loewenstein, Blount, and Bazerman (1999), Prelec, Wernerfelt, and Zettelmeyer (1997), and Tversky, Slovic, and Kahneman (1990).

Overview of the Book

In what follows, we review different topic areas of behavioral economics to place chapters of the book into context. The book is organized so that early chapters discuss basic topics such as decision making under risk and intertemporal choice, while later chapters provide applications of these ideas.

Basic Topics

REFERENCE-DEPENDENCE AND LOSS-AVERSION

In classical consumer theory, preferences among different commodity bundles are assumed to be invariant with respect to an individual's current endowment or consumption. Contrary to this simplifying assumption, diverse forms of evidence point to a dependence of preferences on one's reference point (typically the current endowment). Specifically, people seem to dislike losing commodities from their consumption bundle much more than they like gaining other commodities. This can be expressed graphically as a kink in indifference curves at the current endowment point (Knetsch 1992; Tversky and Kahneman 1991).

In the simplest study showing reference-dependence, Knetsch (1992) endowed some subjects randomly with a mug, while others received a pen.[12] Both groups were allowed to switch their good for the other at a minimal transaction cost, by merely handing it to the experimenter. If preferences are independent of random endowments, the fractions of subjects swapping their mug for a pen and the fraction swapping their pen for a mug should add to roughly one. In fact, 22% of subjects traded. The fact that so few chose to trade implies an exaggerated preference for the good in their endowment, or a distaste for losing what they have.

A seminal demonstration of an "endowment effect" in buying and selling prices was conducted by Kahneman et al. (1990). They endowed half of the subjects in a group with coffee mugs. Those who had mugs were asked the lowest price at which they would sell. Those who did not get mugs were asked how much they would pay. There should be essentially no difference between selling and buying prices. In fact, the median selling price was $5.79 and the median buying price was $2.25, a ratio of more than two: one which has been repeatedly replicated. Although calibrationally entirely implausible, some economists were concerned that the results could be driven by "wealth effects"—those given mugs are wealthier than those not given mugs, and this might make them value mugs more and money less. But in a different study reported in the same paper, the selling prices of one group were compared to the "choosing" prices of another: For a series of money amounts, subjects chose whether they would prefer to have a mug

[12] Note that any possible information value from being given one good rather than the other is minimized because the endowments are random, and subjects knew that half of the others received the good that they didn't have.

or money. The median choosing price was half of the median selling price ($3.50 versus $7.00). Choosers are in *precisely* the same wealth position as sellers—they choose between a mug or money. The only difference is that sellers are "giving up" a mug they "own," whereas choosers are merely giving up the right to have a mug. Any difference between the two groups cannot be attributed to wealth effects.

Kahneman et al.'s work was motivated in part by survey evidence from "contingent valuation" studies that attempt to establish the dollar value of goods that are not routinely traded. Contingent valuation is often used to do government cost-benefit analysis or establish legal penalties from environmental damage. These surveys typically show very large differences between buying prices (e.g., paying to clean up oily beaches) and selling prices (e.g., having to be paid to allow beaches to be ruined). Sayman and Öncüler (1997) summarize 73 data sets that show selling-to-buying ratios ranging from .67 (for raspberry juice) to 20 or higher (for density of trees in a park and health risks).

Loss aversion has already proved to be a useful phenomenon for making sense of field data (see Camerer 2000 and in this volume). Asymmetries in demand elasticities after price increases and decreases (Hardie, Johnson, and Fader 1993), the tendency for New York City cab drivers to quit early after reaching a daily income target, producing surprising upward-sloping labor supply curves (see Camerer et al. 1997 and in this volume), and the large gap between stock and bond returns—the "equity premium" (see Benartzi and Thaler 1995 and in this volume) can all be explained by models in which agents have reference-dependent preferences and take a short planning horizon, so that losses are not integrated against past or future gains.

A particularly conclusive field study by Genegove and Mayer (2001 and in this volume) focuses on the real estate market. (Housing is a huge market—worth $10 trillion at the time of their study, a quarter of the wealth in the United States—and full of interesting opportunities to do behavioral economics.) They find that list prices for condominiums in Boston are strongly affected by the price at which the condominium was purchased. Motivated sellers should, of course, regard the price they paid as a sunk cost and choose a list price that anticipates what the market will pay. But people hate selling their houses at a nominal loss from the purchase price. Sellers' listing prices and subsequent selling behavior reflects this aversion to nominal losses. Odean (1998) finds the same effect of previous purchase price in stock sales.[13]

[13] Though it is harder unambiguously to interpret reference points as loss-aversion in the sense that we are discussing here, they can also serve as social focal points for judging performance. Degeorge, Patel, and Zeckhauser (1999) document an interesting example from corporate finance. Managers whose firms face possible losses (or declines from a previous year's earnings) are very reluctant to report small losses. As a result, the distribution of actual losses and gains shows a very large spike at zero, and hardly any small reported losses (compared to the number of small gains). Wall Street hates to see a small loss. A manager who does not have the skill to shift accounting profits to erase a potential loss (i.e., "has some earnings in his pocket") is considered a poor manager. In this example, the market's aversion to reported losses can serve as a signaling device that tells the markets about managerial ability.

At least three features of endowment effects remain open to empirical discussion. First, do people *anticipate* the endowment effect? The answer seems to be no. Loewenstein and Adler (1995) found that subjects did not anticipate how much their selling prices would increase after they were endowed with mugs.[14] Van Boven, Dunning, and Loewenstein (2000) and Van Boven, Loewenstein, and Dunning (2000) found that agents for buyers also underestimated how much sellers would demand.

Second, Kahneman, Knetsch, and Thaler (1990, p. 1328) note that “there are some cases in which no endowment effect would be expected, such as when goods are purchased for resale rather than for utilization.” However, the boundary of commercial nonattachment has not been carefully mapped. Do art or antique dealers “fall in love” with pieces they buy to resell? What about surrogate mothers who agree to bear a child for a price paid in advance? Evidence on the degree of commercial attachment is mixed. In their housing study, Genesove and Mayer (2001 and in this volume) note that investors who don’t live in their condos exhibit less loss-aversion than owners. A field experiment by List (2003) found that amateur sports paraphernalia collectors who do not trade very often showed an endowment effect, but professional dealers and amateurs who trade a lot did not.[15] An example where attachment seemed important even among experienced traders with high incentives was described by an investment banker who said that his firm combats loss-aversion by forcing a trader periodically to switch his “position” (the portfolio of assets that the trader bought and is blamed or credited for) with the position of another trader. Switching ensures that traders do not make bad trades because of loss-aversion and emotional attachment to their past actions (while keeping the firm’s net position unchanged, since the firm’s total position is unchanged).

Third, it is not clear the degree to which endowment effects are based solely on the current endowment, rather than on past endowments or other reference points. Other reference points, such as social comparison (i.e., the possessions and attainments of other people) and past ownership, may be used to evaluate outcomes. How multiple reference points are integrated is an open question. Strahilevitz and Loewenstein (1998) found that the valuation of objects depended not only on whether an individual was currently endowed with an object, but on the entire past history of ownership—how long the object had been owned or, if it had been lost in the past, how long ago it was lost and how long it was owned before it was lost. These “history-of-ownership effects” were sufficiently strong that choice prices of people who had owned for a long period but who had just lost an object were higher than the selling prices of people who had just acquired the same object.

[14] Failure to anticipate the strength of later loss-aversion is one kind of “projection bias” (Loewenstein, O’Donoghue, and Rabin 1999), in which agents make choices as if their current preferences or emotions will last longer than they actually do.

[15] By revisiting the same traders a year later, List showed that it was trader experience that reduced endowment effects, rather than self-selection (i.e., people who are immune to such effects become dealers.)

If people are sensitive to gains and losses from reference points, the way in which they combine different outcomes can make a big difference. For example, a gain of $150 and a loss of $100 will seem unattractive if they are evaluated separately—if the utility of gains is sufficiently less than the disutility of equal-sized losses, but the gain of $50 that results when the two figures are added up is obviously attractive. Thaler (1980, 1999, and in this volume) suggests that a useful metaphor for describing the rules that govern gain / loss integration is "mental accounting"—people set up mental accounts for outcomes that are psychologically separate, as much as financial accountants lump expenses and revenues into separated accounts to guide managerial attention. Mental accounting stands in opposition to the standard view in economics that "money is fungible"; it predicts, accurately, that people will spend money coming from different sources in different ways (O'Curry 1999), and it has wide-ranging implications for such policy issues as how to promote saving (see Thaler 1994).

A generalization of the notion of mental accounting is the concept of "choice bracketing," which refers to the fashion in which people make decisions narrowly, in a piecemeal fashion, or broadly—i.e., taking account of interdependencies among decisions (Read, Loewenstein, and Rabin 1999). How people bracket choices has far-reaching consequences in diverse areas, including finance (Bernartzi and Thaler 1995, and in this volume), labor supply (Camerer, Babcock, Loewenstein, and Thaler 1997, and in this volume), and intertemporal choice (Frederick, Loewenstein, and O'Donoghue, 2002 and in this volume). For example, when making many separate choices among goods, people tend to choose more diversity when the choices are bracketed broadly than when they are bracketed narrowly. This was first demonstrated by Simonson (1990), who gave students their choice of one of six snacks during each of three successive weekly class meetings. Some students chose all three snacks in the first week, although they didn't receive their chosen snack until the appointed time, and others chose each snack on the day that they were to receive it (narrow bracketing; sequential choice). Under broad bracketing, fully 64% chose a different snack for each week, as opposed to only 9% under narrow bracketing. Follow-up studies demonstrated similar phenomena in the field (e.g., in purchases of yogurt; Simonson and Winer 1992).

Bracketing also has implications for risk-taking. When people face repeated risk decisions, evaluating those decisions in combination can make them appear less risky than if they are evaluated one at a time. Consequently, a decision maker who refuses a single gamble may nonetheless accept two or more identical ones. By assuming that people care only about their overall level of wealth, expected-utility theory implicitly assumes broad bracketing of risky decisions. However, Rabin (2000) points out the absurd implication that follows from this assumption (combined with the assumption that risk-aversion stems from the curvature of the utility function): A reasonable amount of aversion toward risk in small gambles implies a dramatic aversion to reduction in overall wealth. For example, a person who will turn down a coin flip to win $11 and lose $10 at all wealth levels must also turn down a coin flip in which she can lose $100, *no matter how large the* *possible gain is*.[16] Rabin's proof is a mathematical demonstration that people who are averse to small risks are probably not integrating all their wealth into one source when they think about small gambles.

PREFERENCES OVER RISKY AND UNCERTAIN OUTCOMES

The expected-utility (EU) hypothesis posits that the utility of a risky distribution of outcomes (say, monetary payoffs) is a probability-weighted average of the outcome utilities. This hypothesis is normatively appealing because it follows logically from apparently reasonable axioms, most notably the independence (or "cancellation") axiom. The independence axiom says that if you are comparing two gambles, you should cancel events that lead to the same consequence with the same probability; your choice should be independent of those equally likely common consequences. Expected utility also simplifies matters because a person's taste for risky money distributions can be fully captured by the shape of the utility function for money.

Many studies document predictive failures of expected utility in simple situations in which subjects can earn substantial sums of money from their choices.[17] Starmer's (2000) contribution to this volume reviews most of these studies, as well as the many theories that have been proposed to account for the evidence (see also Camerer 1989b, 1992; Hey 1997; Quiggin 1993). Some of these new theories alter the way in which probabilities are weighted but preserve a "betweenness" property that says that if A is preferred to B, then any probabilistic gamble between them must be preferred to B but dispreferred to A (i.e., the gambles lie "between" A and B in preference). Other new theories suggest that probability weights are "rank-dependent"—outcomes are first ranked, then their probabilities are weighted in a way that is sensitive to how they rank within the gamble that is being considered. One mathematical way to do this is transform

[16] The intuition behind Rabin's striking result is this: In expected-utility theory, rejecting a (+\$11, −\$10) coin flip at wealth level W implies that the utility increase from the \$11 gain is smaller than the total utility decrease from the \$10 loss, meaning that the marginal utility of each dollar gained is at most 10/11 of the marginal utility of each dollar lost. By concavity, this means that the marginal utility of the W + 11th dollar is at most 10/11 the marginal utility of the W − 10th dollar—a sharp 10% drop in marginal utility for small change in overall wealth of \$21. When the curvature of the utility function does not change unrealistically over ranges of wealth levels, this means the marginal utility plummets quickly as wealth increases—the marginal utility of the W + \$32 dollar (=W + 11 + 21) can be at most (10/11)(10/11), which is around 5/6 of the marginal utility of the W − 10th dollar. Every \$21 decrease in wealth yields another 10% decline in marginal utility. This suggests, mathematically, that implying a person's value for a dollar if he were \$500 or \$1,000 wealthier would be tiny compared to how much he values dollars that he might lose in a bet. So if a person's attitude toward gambles really came from the utility-of-wealth function, even incredibly large gains in wealth would not tempt her to risk \$50 or \$100 losses, if she really dislikes losing \$10 more than she likes gaining \$11 at every level of wealth.

[17] Some of the earlier studies were done with hypothetical payoffs, leading to speculation that the rejection of EU would not persist with real stakes. Dozens of recent studies show that, in fact, paying real money instead of making outcomes hypothetical either fails to eliminate EU rejections or *strengthens* the rejections of EU (because sharper results that come from greater incentive imply that rejections are more statistically significant; Harless and Camerer 1994).

the *cumulative* probabilities of outcomes (i.e., the chance that you will win X or less) nonlinearly and weigh outcome utilities by the differences of those weighted cumulative probabilities.[18] The best-known theory of this sort is cumulative prospect theory (Tversky and Kahneman 1992).

There are three clear conclusions from the experimental research (Harless and Camerer 1994). One is that of the two new classes of theories that allow more general functional forms than expected utility, the new rank-dependent theories fit the data better than the new betweenness class theories. A second conclusion is that the statistical evidence against EU is so overwhelming that it is pointless to run more studies testing EU against alternative theories (as opposed to comparing theories with one another). The third conclusion is that EU fits worst when the two gambles being compared have different sets of possible outcomes (or "support"). Technically, this property occurs when one gamble has a unique outcome. The fact that EU does most poorly for these comparisons implies that nonlinear weighting of low probabilities is probably a major source of EU violations. Put differently, EU is like Newtonian mechanics, which is useful for objects traveling at low velocities but mispredicts at high speeds. Linear probability weighting in EU works reasonably well except when outcome probabilities are very low or high. But low-probability events are important in the economy, in the form of "gambles" with positive skewness (lottery tickets, and also risky business ventures in biotech and pharmaceuticals), and catastrophic events that require large insurance industries.

Prospect theory (Kahneman and Tversky 1979) explains experimental choices more accurately than EU because it gets the psychophysics of judgment and choice right. It consists of two main components: a probability weighting function, and a "value function" that replaces the utility function of EU. The weighting function $\pi(p)$ combines two elements: (1) The level of probability weight is a way of expressing risk tastes (if you hate to gamble, you place low weight on any chance of winning anything); and (2) the curvature in $\pi(p)$ captures how sensitive people are to differences in probabilities. If people are more sensitive in the neighborhoods of possibility and certainty—i.e., changes in probability near zero and 1—than to intermediate gradations, then their $\pi(p)$ curve will overweight low probabilities and underweight high ones.

The value function reflects the insight, first articulated by Markowitz (1952), that the utility of an outcome depends not on the absolute level of wealth that results but on whether the outcome is a gain or a loss. Prospect theory also assumes reflection of risk-preferences at the reference point: People are typically averse to risky spreading of possible money gains, but will take gambles where they could

[18] A technical motivation for "rank dependent" theories—ranking outcomes, then weighting their probabilities—is that when separate probabilities are weighted, it is easy to construct examples in which people will violate dominance by choosing a "dominated" gamble A, which has a lower chance of winning at each possible outcome amount, compared to the higher chance of winning the same outcome amount for a dominant gamble B. If people rarely choose such dominated gambles, they are acting as if they are weighting the *differences* in cumulated probabilities, which is the essence of the rank-dependent approaches.

lose big or break even rather than accept a sure loss. Prospect theory also assumes "loss-aversion": The disutility of a loss of x is worse than the utility of an equal-sized gain of x.

Expected utility is restricted to gambles with known outcome probabilities. The more typical situation in the world is "uncertainty," or unknown (subjective, or personal) probability. Savage (1954) proposed a subjective expected utility (SEU) theory in which choices over gambles would reveal subjective probabilities of states, as well as utilities for outcomes. Ellsberg (1961) quickly pointed out that in Savage's framework, subjective probabilities are slaves to two masters—they are used as decision weights applied to utilities and they are expressions of likelihood. As a result, there is no way to express the possibility that, because a situation is "ambiguous," one is reluctant to put much decision weight on *any* outcome. Ellsberg demonstrated this problem in his famous paradox: Many people prefer to bet on black drawn from an urn with 50 black and 50 red balls, rather than bet on black drawn from an urn with 100 balls of unknown black and red composition, and similarly for red (they just don't want to bet on the unknown urn). There is no way for the two sets of red and black subjective probabilities from each urn both to add to one (as subjective probabilities require), and still express the distaste for betting neither color in the face of ambiguity.

Many theories have been proposed to generalize SEU to allow for ambiguity-aversion (see Camerer and Weber [1992] for a review). One approach, first proposed by Ellsberg, is to let probabilities be *sets* rather than specific numbers, and to assume that choices over gambles reveal whether or not people pessimistically believe the worst probabilities are the right ones. Another approach is to assume that decision weights are nonadditive. For example, the weights on red and black in the Ellsberg unknown urn could both be .4; the missing weight of .2 is a kind of "reserved belief" that expresses how much the person dislikes betting when she knows that important information is missing.

Compared to non-EU theories, relatively little empirical work and applications have been done with these uncertainty-aversion theories so far. Uncertainty-aversion might explain phenomena like voting "roll-off" (when a voter, once in the voting booth, refuses to vote on obscure elections in which their vote is most likely to prove pivotal; Ghirardato and Katz 2000), incomplete contracts (Mukherji 1998) and "home country bias" in investing: People in every country overinvest in the country they are most familiar with—their own. (Finnish people invest in firms closer to their own town; see Grinblatt and Keloharju 2001.)

In asset pricing, ambiguity-aversion can imply that asset prices satisfy a pair of Euler inequalities, rather than an Euler equation, which permits asset prices to be more volatile than in standard theory (Epstein and Wang 1994). Hansen, Sargent, and Tallarini (1999) have applied related concepts of "robust control" to macroeconomic fluctuations. Finally, uncertainty-averse agents will value information even if it does not change the decisions that they are likely to make after becoming better informed (simply because information can make nonadditive decision weights closer to additive and can make agents "feel better" about their decision). This effect may explain demand for information in settings like medicine or personal finance, where new information usually does not change choices but relieves anxiety people have from knowing that there is something they could know but do not (Asch, Patton, and Hershey 1990).

INTERTEMPORAL CHOICE

The discounted-utility (DU) model assumes that people have instantaneous utilities from their experiences each moment, and that they choose options that maximize the present discounted sum of these instantaneous utilities. Typically it is assumed that instantaneous utility each period depends solely on consumption in that period, and that the utilities from streams of consumption are discounted exponentially, applying the same discount rate in each period. Samuelson (1937) proposed this particular functional form because it was simple and similar to present value calculations applicable to financial flows. But in the article in which he proposed the DU model, he repeatedly drew attention to its psychological implausibility.[19] Decades of empirical research substantiated his doubts (see Loewenstein and Prelec 1992, and Frederick, Loewenstein and O'Donoghue, 2002, and in this volume).

It is useful to separate studies dealing with intertemporal choice into those that focus on phenomena that can be explained on the basis of the discount function and those that can be explained on the basis of the utility function. The following two subsections cover these points.

TIME DISCOUNTING

A central issue in economics is how agents trade off costs and benefits that occur at different points in time. The standard assumption is that people weight future utilities by an exponentially declining discount factor $d(t) = \delta^t$, where $1 > \delta > 0$. Note that the discount factor δ is often expressed as $1/(1 + r)$, where r is a discount *rate*.

However, a simple hyperbolic time discounting function of $d(t) = 1/(1 + kt)$ tends to fit experimental data better than exponential discounting. The early evidence on discounting came from studies showing that animals exhibit much large discounting when comparing immediate rewards and rewards delayed t periods, compared to the trade-off between rewards k and $k + t$ periods in the future. Thaler (1981) was the first to test empirically the constancy of discounting with human subjects. He told subjects to imagine that they had won some money in a lottery held by their bank. They could take the money now or earn interest and wait until later. They were asked how much they would require to make waiting just as attractive as getting the money immediately. Thaler then estimated implicit (per-period) discount rates for different money amounts and time delays under the assumption that subjects had linear utility functions. Discount rates declined linearly with the duration of the time delay. Later studies replicated the basic finding that discount rates fall with duration (Benzion, Rapoport, and Yagil 1989;

[19] The notion of discounting utility at a fixed rate was first mentioned, in passing, in an article by Ramsey (1928) on intergenerational saving.

Holcomb and Nelson, 1992). The most striking effect is an "immediacy effect" (Prelec and Loewenstein 1991): discounting is dramatic when one delays consumption that would otherwise be immediate.

Declining discount rates have also been observed in experimental studies involving real money outcomes. Horowitz (1992) tested the constancy of discounting by auctioning "bonds" in a Vickrey (highest-rejected-bid) auction. The amount bid for a bond represented how much a subject was willing to give up at the time of the auction for certain future payoffs. Discount rates again decreased as the horizon grew longer. Pender (1996) conducted a study in which Indian farmers made several choices between amounts of rice that would be delivered either sooner or later. Fixing the earlier rice ration and varying the amount of rice delivered later gives an estimate of the discount rate. To avoid immediacy effects, none of the choices was delivered immediately. Per-period discount rates decline with the increasing horizon: the mean estimated discount rate was .46 for 7 months and .33 for 5 years.

Hyperbolic time discounting implies that people will make relatively farsighted decisions when planning in advance—when all costs and benefits will occur in the future—but will make relatively shortsighted decisions when some costs or benefits are immediate. The systematic changes in decisions produced by hyperbolic time discounting create a time-inconsistency in intertemporal choice not present in the exponential model. An agent who discounts utilities exponentially would, if faced with the same choice and the same information, make the same decision prospectively as he would when the time for a decision actually arrived. In contrast, somebody with time-inconsistent hyperbolic discounting will wish prospectively that in the future he would take farsighted actions; but when the future arrives he will behave against his earlier wishes, pursuing immediate gratification rather than long-run well-being.

Strotz (1955) first recognized the planning problem for economic agents who would like to behave in an intertemporally consistent fashion, and discussed the important ramifications of hyperbolic time discounting for intertemporal choice. Most big decisions—regarding, e.g., savings, educational investments, labor supply, health and diet, crime and drug use—have costs and benefits that occur at different points in time. Many authors such as Thaler (1981), Thaler and Shefrin (1981), and Schelling (1978) discussed the issues of self-control and stressed their importance for economics. Laibson (1997) accelerated the incorporation of these issues into economics by adopting a "quasi-hyperbolic" time discounting function (first proposed by Phelps and Pollak [1968] to model intergenerational utility). The quasi-hyperbolic form approximates the hyperbolic function with two parameters, β and δ, in which the weight on current utility is 1 and the weight on period-t instantaneous utility is $\beta\delta^t$ for $t > 0$. The parameter β measures the immediacy effect: if $\beta = 1$ the model reduces to standard exponential discounting. When delayed rewards are being compared, the immediacy premium β divides out so that the ratio of discounted utilities is solely determined by δ^t (consistent with the observations of Benzion, Rapoport, and Yagil 1989).

Thus, quasi-hyperbolic time discounting is basically standard exponential time discounting plus an immediacy effect; a person discounts delays in gratification equally at all moments except the current one—caring differently about well-being now versus later. This functional form provides one simple and powerful model of the taste for immediate gratification.

In his 1997 paper, reprinted in chapter 15 of this volume, Laibson applies the quasi-hyperbolic model to a model of lifetime consumption-savings decisions. He emphasizes the role that the partial illiquidity of an asset plays in helping consumers constrain their own future consumption. If people can withdraw money immediately from their assets, as they can with simple savings or checking accounts, they have no way to control their temptation to overconsume. Assets that are less liquid, despite their costly lack of flexibility or even lower yield, may be used as a commitment device for those consumers who at least partially understand their tendency to overconsume. In this paper and others (including the more recent papers coauthored by Laibson, Repetto, and Tobacman [1998]), it has been demonstrated how quasi-hyperbolic discounting potentially provides a better account than does conventional exponential discounting of various savings and consumption phenomena, such as different marginal propensities to consume out of different forms of savings, and the ways that financial innovation (typically in the form of increased liquidity) may lead to damaging decreases in savings.

An important question in modeling self-control is whether agents are aware of their self-control problem ("sophisticated") or not ("naïve"). The work in macroeconomics described above assumes agents are sophisticated, but have some commitment technologies to limit how much the current self can keep the future self from overspending.[20] However, there are certainly many times in which people are partially unaware of their own future misbehavior, and hence overly optimistic that they will behave in the future the way in which that their "current self" would like them to. O'Donoghue and Rabin (1999 and in this volume; cf. Akerlof 1991) show how awareness of self-control problems can powerfully moderate the behavioral consequences of quasi-hyperbolic discounting.

Naïveté typically makes damage from poor self-control worse. For example, severe procrastination is a creation of overoptimism: One can put off doing a task repeatedly if the perceived costs of delay are small—"I'll do it tomorrow, so there is little loss from not doing it today"—and hence accumulate huge delay costs from postponing the task many times. A sophisticated agent aware of his procrastination will realize that if he puts if off, he will only have to do the task in the future, and hence will do it immediately. However, in some cases, being sophisticated about one's self-control problem can *exacerbate* yielding to temptation. If you are aware of your tendency to yield to a temptation in the future, you may conclude that you might as well yield now; if you naïvely think you will resist temptation for longer in the future, that may motivate you to think it is worthwhile resisting temptation now. More recently, O'Donoghue and Rabin (2001) have developed a model of "partial naïveté" that permits a whole continuum of degree of awareness, and many other papers on quasi-hyperbolic discounting have begun to

[20] Ariely and Wertenbroch (in press) report similar self-commitment—deadline-setting—in an experiment.

clarify which results come from the quasi-hyperbolic preferences per se and which come from assumptions about self-awareness of those preferences.

Many of the most striking ways in which the classical DU model appears to fail stem not from time discounting but from characteristics of the utility function. Numerous survey studies (Benzion et al. 1989; Loewenstein 1988; Thaler 1981) have shown that gains and losses of different absolute magnitudes are discounted differently. Thaler's (1981) subjects were indifferent toward receiving $15 immediately and receiving $60 in a year (a ratio of .25) and also between $250 immediately and $350 in a year (a ratio of .71). Loewenstein and Prelec (1992) replicated these "magnitude effects," and also showed that estimated discount rates for losses tend to be lower than those for gains. Again, these effects are inconsistent with DU. A third anomaly is that people dislike "temporal losses"—delays in consumption—much more than they like speeding up consumption (Loewenstein 1988).

None of these effects can be explained by DU, but they are consistent with a model proposed by Loewenstein and Prelec (1992). This model departs from DU in two major ways. First, as discussed in the previous subsection, it incorporates a hyperbolic discount function. Second, it incorporates a utility function with special curvature properties that is defined over gains and losses rather than final levels of consumption. Most analyses of intertemporal choice assume that people integrate new consumption with planned consumption. While such integration is normatively appealing, it is computationally infeasible and, perhaps for this reason, descriptively inaccurate. When people make decisions about new sequences of payments or consumption, they tend to evaluate them in isolation—e.g., treating negative outcomes as losses rather than as reductions to their existing money flows or consumption plans. No model that assumes integration can explain the anomalies just discussed.

Such anomalies are sometimes mislabeled as discounting effects. It is said that people "discount" small outcomes more than large ones, gains more than losses, and that they exhibit greater time discounting for delay than for speedup. Such statements are misleading. In fact, all of these effects are consistent with stable, uniform, time discounting once one measures discount rates with a more realistic utility function. The inconsistencies arise from misspecification of the utility function, not from differential time discounting of different types of outcomes.

Another anomaly is apparent *negative* time discounting. If people like savoring pleasant future activities they may postpone them to prolong the pleasure (and they may get painful activities over with quickly to avoid dread). For example, Loewenstein (1987) elicited money valuations of several outcomes that included a "kiss from the movie star of your choice," and "a nonlethal 110 volt electric shock" occurring at different points in time. The average subject paid the most to delay the kiss three days and was eager to get the shock over with as quickly as possible (see also Carson and Horowitz 1990; MacKeigan et al. 1993). In a standard DU model, these patterns can be explained only by discount factors that are greater than one (or discount *rates* that are negative). However, Loewenstein (1987) showed that these effects can be explained by a model with positive time discounting, in which people derive utility (both positive and negative) from anticipation of future consumption.

A closely related set of anomalies involves sequences of outcomes. Until recently, most experimental research on intertemporal choice involved single outcomes received at a single point in time. The focus was on measuring the correct form of the discount function and it was assumed that once this was determined, the value of a sequence of outcomes could be arrived at by simply adding up the present values of its component parts. The sign and magnitude effects and the delay / speedup asymmetry focused attention on the form of the utility function that applies to intertemporal choice, but retained the assumption of additivity across periods. Because they involved only single outcomes, these phenomena shed no light on the validity of the various independence assumptions that involve multiple time periods.

Research conducted during the past decade, however, has begun to examine preferences toward sequences of outcomes and has found quite consistently that they do not follow in a simple fashion from preferences for their component parts (Loewenstein and Prelec 1993). People care about the "gestalt," or overall pattern of a sequence, in a way that violates independence.

A number of recent studies have shown that people generally favor sequences that improve over time. Loewenstein and Sicherman (1991) and Frank and Hutchens (1993 and this volume), for example, found that a majority of subjects prefer an increasing wage profile to a declining or flat one, for an otherwise identical job. Preference for improvement appears to be driven in part by savoring and dread (Loewenstein 1987), and in part by adaptation and loss-aversion. Savoring and dread contribute to preference for improvement because, for gains, improving sequences allows decision makers to savor the best outcome until the end of the sequence. With losses, getting undesirable outcomes over with quickly eliminates dread. Adaptation leads to a preference for improving sequences because people tend to adapt to ongoing stimuli over time and to evaluate new stimuli relative to their adaptation level (Helson, 1964), which means that people are sensitive to *change*. Adaptation favors increasing sequences, which provide a series of positive changes—i.e., *gains*—over decreasing sequences, which provide a series of negative changes—i.e., *losses*. Loss-aversion intensifies the preference for improvement over deterioration (Kahneman and Tversky 1979).

The idea that adaptation and loss-aversion contribute to the preference for sequences, over and above the effects of savoring and dread, was suggested by a study conducted by Loewenstein and Prelec (1993). They asked subjects first to state a preference between a fancy French restaurant dinner for two either on Saturday in one month or Saturday in two months. Eighty percent preferred the more immediate dinner. Later, the same respondents were asked whether they would prefer the sequence fancy French this month and mediocre Greek next month, or mediocre Greek this month and fancy French next month. When the choice was expressed as one between sequences, a majority of respondents shifted in favor of preferring the improving sequence—which delayed the French dinner for two months. The same pattern was observed when the mediocre Greek restaurant was replaced by "eat at home," making it even more transparent that the sequence frame was truly changing people's preferences. The conclusion of this research is that, as in visual perception, people have a "gestalt" notion of an ideal distribution of outcomes in time, which includes interactions across time periods that violate simple separability axioms.

FAIRNESS AND SOCIAL PREFERENCES

The assumption that people maximize their own wealth and other personal material goals (hereafter, just "self-interest") is a widely correct simplification that is often useful in economics. However, people may sometimes choose to "spend" their wealth to punish others who have harmed them, reward those who have helped, or to make outcomes fairer. Just as understanding demand for goods requires specific utility functions, the key to understanding this sort of social preference is a parsimonious specification of "social utility," which can explain many types of data with a single function.

An experimental game that has proved to be a useful workhorse for identifying departures from self-interest is the "ultimatum" game, first studied by Güth et al. (1982). In an ultimatum game, a proposer has an amount of money, typically about $10, from which he must propose a division between himself and a responder. (The players are anonymous and will never see each other again.) If the responder accepts the offered split, they both get paid and the game ends. If she rejects the offer, they get nothing and the game ends. In studies in more than 20 countries, the vast majority of proposers offer between a third and a half of the total, and responders reject offers of less than a fifth of the total about half of the time. A responder who rejects an offer is spending money to punish somebody who has behaved unfairly.

A "trust" game can be used to explore the opposite pattern, "positive reciprocity." Positive reciprocity means that players are disposed to reward those who have helped them, even at a cost to themselves. In a typical trust game, one player has a pot of money, again typically around $10, from which he can choose to keep some amount for himself, and to invest the remaining amount X, between $0 and $10, and their investment is tripled. A trustee then takes the amount 3X, keeps as much as she wants, and returns Y. In standard theory terms, the investor-trustee contract is incomplete and the investor should fear trustee moral hazard. Self-interested trustees will keep everything ($Y = 0$) and self-interested investors who anticipate this will invest nothing ($X = 0$). In fact, in most experiments investors invest about half and trustees pay back a little less than the investment. Y varies positively with X, as if trustees feel an obligation to repay trust.

The first to attempt to model these sorts of patterns was Rabin (1993, and this volume). Fixing player A's likely choice, player B's choice determines A's payoff. From A's point of view, B's choice can be either kind (gives A a lot) or mean (gives A very little). This enables A to form a numerical judgment about B's kindness, which is either negative or positive (zero represents kindness-neutrality). Similarly, A's action is either kind or mean toward B. In Rabin's approach, people earn a utility from the payoff in the game and a utility from the product of their kindness and the kindness of the other player. Multiplying the two kindness terms generates both negative and positive reciprocity, or a desire for emotional coordination: If B is positively kind, A *prefers* to be kind too; but if B is mean (negative kindness), then A prefers to be mean. Rabin then uses concepts from game theory to derive consequences for equilibrium, assuming people have fairness-adjusted utilities.[21]

Besides explaining some classic findings, Rabin's kindness-product approach makes fresh predictions: For example, in a prisoner's dilemma (PD), mutual cooperation can be a "fairness equilibrium." (Cooperating is nice; therefore, reciprocating anticipated cooperation is mutually nice and hence utility-maximizing.) But if player A is *forced* to cooperate, then player A is not being kind and player B feels no need to behave kindly. So player B should defect in the "involuntary" PD.

Other approaches posit a social utility function that combines one's own payoff with her relative share of earnings, or the difference between her payoffs and the payoffs of others. One example is Fehr and Schmidt (1999 and in this volume), who use the function $u_i(x_1, x_2, \ldots, x_n) = x_i - \alpha\Sigma_k[x_k - x_i]_0/(n-1) - \beta\Sigma_k[x_i - x_k]_0/(n-1)$, where $[x]_0$ is x if $x > 0$ and 0 otherwise. The coefficient α is the weight on envy or disadvantageous inequality (when $x_k > x_i$), and β is the weight on guilt or advantageous inequality ($x_i > x_k$). This inequality-aversion approach matches ultimatum rejections because an offer of \$2 from a \$10 pie, say, has utility $2 - (8-2)\alpha$ while rejecting yields 0. Players who are sufficiently envious ($\alpha > 1/3$) will reject such offers. Inequality-aversion also mimics the effect of positive reciprocity because players with positive values of   will feel sheepish about earning more money than others do; so they will repay trust and feel bad about defecting in PDs and free-riding in public goods contribution games. Bolton and Oeckenfels (2000) propose a similar model.

Charness and Rabin (forthcoming) propose a "Rawlsitarian" model that integrates three factors—one's own payoff, and a weighted average of the lowest payoff anyone gets (à la Rawls) and the sum of everyone's payoff (utilitarian). This utility function explains new results from three-person games that are not explained by the inequality-aversion forms, and from a large sample of two-person games where the inequality-aversion approaches often predict poorly.

The key point is that careful experimental study of simple games in which social preferences play a role (like ultimatum and trust) has yielded tremendous regularity. The regularity has, in turn, inspired different theories that map payoffs to all players into each player's utility, in a parsimonious way. Several recent papers compare the predictions of different models (see Camerer 2003, chap. 2). The results show that some form of the intentionality incorporated in Rabin (1993 and in this volume; players care about whether another player *meant* to harm them or help them), combined with inequality-aversion or Rawlsitarian mixing will explain a good amount of data. Models like these also make new predictions and should be useful in microeconomics applications as well.

Kahneman, Knetsch, and Thaler (1986 and in this volume) studied consumer perceptions of fairness using phone surveys. They asked people about how fair

[21] He used the theory of psychological games, in which a player's utilities for outcomes can depend on their beliefs (Geanakopolos, Pearce, and Stacchetti 1989). For example, a person may take pleasure in being surprised by receiving a gift, aside from the gift's direct utility.

they considered different types of firm behavior to be. In a typical question, they asked people whether a hardware store that raised the price of a snow shovel after a snowstorm was behaving fairly or not. (People thought the store was unfair.) Their results can be neatly summarized by a “dual-entitlement” hypothesis: Previous transactions establish a reference level of consumer surplus and producer profit. Both sides are “entitled” to these levels of profit, and so price changes that threaten the entitlement are considered unfair.

Raising snow-shovel prices after a snowstorm, for example, reduces consumer surplus and is considered unfair. But when the cost of a firm’s inputs rises, subjects said it was fair to raise prices—because *not* raising prices would reduce the firm’s profit (compared to the reference profit). The Kahneman et al. framework has found surprisingly little application, despite the everyday observation that firms do not change prices and wages as frequently as standard theory suggests. For example, when the fourth Harry Potter book was released in summer 2000, most stores were allocated a small number of books that were sold in advance. Why not raise prices, or auction the books off? Everyday folks, like the subjects in Kahneman et al.’s surveys, find actions that exploit excess demand to be outrageous. Concerned about customer goodwill, firms limit such price increases.

An open question is whether consumers are really willing to express outrage at unfairness by boycotts and other real sacrifices (Engelmann and Tyran [2002] find that boycotts are common in the lab). A little threat of boycott also may go a long way toward disciplining firms. (In the ultimatum game, for example, many subjects *do* accept low offers; but the fraction that reject such offers is high enough that it pays for proposers to offer almost half.) Furthermore, even if consumer boycotts rarely work, offended consumers are often able to affect firm behavior by galvanizing media attention or provoking legislation. For example, “scalping” tickets for popular sports and entertainment events (reselling them at a large premium over the printed ticket price) is constrained by law in most states. Some states have “anti-gouging” laws penalizing sellers who take advantage of shortages of water, fuel, and other necessities by raising prices after natural disasters. A few years ago, responding to public anger at rising CEO salaries when the economy was being restructured through downsizing and when many workers lost their jobs, Congress passed a law prohibiting firms from deducting a CEO salary, for tax purposes, beyond $1 million a year (Rose and Wolfram 2000). Explaining where these laws and regulations come from is one example of how behavioral economics might be used to expand the scope of law and economics (see Sunstein 2000).

BEHAVIORAL GAME THEORY

Game theory has rapidly become an important foundation for many areas of economic theory, such as bargaining in decentralized markets, contracting and organizational structure, as well as political economy (e.g., candidates choosing platforms and congressional behavior). The descriptive accuracy of game theory in these applications can be questioned because equilibrium predictions often assume sophisticated strategic reasoning, and direct field tests are difficult. As a result, there have been many experiments that test game-theoretic predictions.

"Behavioral game theory" uses this experimental evidence and psychological intuition to generalize the standard assumptions of game theory in a parsimonious way. Some of the experimental evidence, and its relation to standard ideas in game theory, is reviewed by Crawford (1997 and in this volume). Newer data and theories that explain them are reviewed briefly by Goeree and Holt (1999) and at length by Camerer (in this volume).

One component of behavioral game theory is a theory of social preferences for allocations of money to oneself and others (discussed above). Another component is a theory of how people choose in one-shot games or in the first period of a repeated game. A simple example is the "p-beauty contest game": Players choose numbers in [0,100] and the player whose number is closest in absolute value to p times the average wins a fixed prize. (The game is named after a well-known passage in which Keynes compared the stock market to a "beauty contest" in which investors care only about what stocks others think are "beautiful.") There are many experimental studies for $p = 2/3$. In this game the unique Nash equilibrium is zero. Since players want to choose 2/3 of the average number, if they think that others will choose 50, for example, they will choose 33. But if they think that others will use the same reasoning and hence choose 33, they will want to choose 22. Nash equilibrium requires this process to continue until players beliefs' and choices match. The process stops, mathematically, only when $x = (2/3)x$, yielding an equilibrium of zero.

In fact, subjects in p-beauty contest experiments seem to use only one or two steps of iterated reasoning: Most subjects best respond to the belief that others choose randomly (step 1), choosing 33, or best respond to step-1 choices (step 2), choosing 22. (This result has been replicated with many subject pools, including Caltech undergraduates with median math SAT scores of 800 and corporate CEOs; see Camerer, Ho, and Chong 2003.)

Experiments like these show that the mutual consistency assumed in Nash equilibrium—players correctly anticipate what others will do—is implausible the first time players face a game, and so there is room for a theory that is descriptively more accurate. A plausible theory of this behavior is that players use a distribution of decision rules, like the steps that lead to 33 and 22, or other decision rules (Stahl and Wilson 1995; Costa-Gomes, Crawford, and Broseta 2001). Camerer, Ho, and Chong (2003) propose a one-parameter cognitive hierarchy (CH) model in which the frequency of players using higher and higher steps of thinking is given by a one-parameter Poisson distribution). If the mean number of thinking steps is specified in advance (1.5 is a reasonable estimate), this theory has zero free parameters, is just as precise as Nash equilibrium (sometimes more precise), and always fits experimental data better (or equally well).

A less behavioral alternative that maintains the Nash assumption of mutual consistency of beliefs and choices is a stochastic or "quantal-response" equilibrium (QRE; see Goeree and Holt [1999]; McKelvey and Palfrey [1995, 1998]; cf. Weizsacker, in press). In a QRE, players form beliefs about what others will do, and calculate the expected payoffs of different strategies, but they do not always choose the best response with the highest expected payoff (as in Nash equilibrium).

Instead, strategies are chosen according to a statistical rule in which better responses are chosen more often. QRE is appealing because it is a minimal (one-parameter) generalization of Nash equilibrium, which avoids many of the technical difficulties of Nash and fits data better.[22]

A third component of behavioral game theory is a model of learning. Game theory is one area of economics in which serious attention has been paid to the process by which an equilibrium comes about. A popular approach is to study the evolution of a population (abstracting from details of how different agents in the population learn). Other studies posit learning by individual agents, based on their own experience or on imitation (Schlag 1998). Many learning theories have been proposed and carefully tested with experimental data. Theories about population evolution never predict as well as theories of individual learning (though they are useful for other purposes). In reinforcement theories, only chosen strategies get reinforced by their outcomes (Roth et al. 2000). In belief-learning theories, players change their guesses about what other players will do, based on what they have seen, and choose strategies that have high expected payoffs, given those updated guesses (Fudenberg and Levine 1998). In the hybrid "experience weighted attraction" (EWA) theory of Camerer and Ho (1999), players respond weakly to "foregone payoffs" from unchosen strategies and more strongly to payoffs that they actually receive (as if underweighting "opportunity costs"; see Thaler 1999 and in this volume). Reinforcement and "fictitious play" theories of belief learning are boundary cases of the EWA theory. In many games (e.g., those with mixed-strategy equilibria), these theories are about equally accurate and are better than equilibrium theories. However, EWA is more robust in the sense that it predicts accurately in games where belief and reinforcement theories don't predict well (see Camerer, Ho, and Chong 2002).

Some next steps are to explore theoretical implications of the theories that fit data well and to understand learning in very complex environments. The most important direction is application to field settings. Two interesting examples are the industrial structure in the Marseilles fish market (Weisbuch, Kirman, and Herreiner 2000) and a large sample (130,000) of consumer supermarket purchases (Ho and Chong, 2003).

Applications

MACROECONOMICS AND SAVING

Many concepts in macroeconomics probably have a behavioral underpinning that could be elucidated by research in psychology. For example, it is common to assume that prices and wages are rigid (in nominal terms), which has important

[22] A classic problem is how players in a dynamic game update their beliefs off the equilibrium path, when a move that (in equilibrium) has zero probability occurs. Bayes's rule cannot be used because P(event) = 0, so any conditional probability P(state | event) divides by zero. QRE sidesteps this problem because stochastic responses ensure that all events have positive probability. This solution is much like the "trembles" proposed by Selten and like subsequent refinements, except that the tremble probabilities are endogeneous.

implications for macroeconomic behavior. Rigidities are attributed to a vague exogeneous force like "menu costs," shorthand for some unspecified process that creates rigidity. Behavioral economics suggests ideas as to where rigidity comes from. Loss-aversion among consumers and workers, perhaps inflamed by workers' concern for fairness, can cause nominal rigidity but are rarely discussed in the modern literature (though see Bewley 1998; Blinder et al. 1998).

An important model in macroeconomics is the life-cycle model of savings (or permanent income hypothesis). This theory assumes that people make a guess about their lifetime earnings profile and plan their savings and consumption to smooth consumption across their lives. The theory is normatively appealing if consumption in each period has diminishing marginal utility, and if preferences for consumptions streams are time-separable (i.e., overall utility is the sum of the discounted utility of consumption in each separate period). The theory also assumes that people lump together different types of income when they guess how much money they'll have (i.e., different sources of wealth are fungible).

Shefrin and Thaler (1992 and in this volume) present a "behavioral life cycle" theory of savings in which different sources of income are kept track of in different mental accounts. Mental accounts can reflect natural perceptual or cognitive divisions. For example, it is possible to add up your paycheck and the dollar value of your frequent flyer miles, but it is simply unnatural (and a little arbitrary) to do so, like measuring the capacity of your refrigerator by how many calories it holds. Mental accounts can also be bright-line devices to avoid temptation: Allow yourself to head to Vegas after cashing an IRS refund check, but not after raiding the childrens' college fund or taking out a housing equity loan. Shefrin and Thaler show that plausible assumptions about mental accounting for wealth predict important deviations from life-cycle savings theory. For example, the measured marginal propensities to consume (MPC) an extra dollar of income from different income categories are very different. The MPC from housing equity is extremely low (people don't see their house as a pile of cash). On the other hand, the MPC from windfall gains is substantial and often close to 1 (the MPC from one-time tax cuts is around 1/3–2/3).

It is important to note that many key implications of the life-cycle hypothesis have *never* been well supported empirically (e.g., consumption is far more closely related to current income than it should be, according to theory). Admittedly, since empirical tests of the life-cycle model involve many auxiliary assumptions, there are many possible culprits if the theory's predictions are not corroborated. Predictions can be improved by introducing utility functions with "habit formation," in which utility in a current period depends on the reference point of previous consumption, and by more carefully accounting for uncertainty about future income (see Carroll 2000). Mental accounting is only one of several behavioral approaches that may prove useful.

An important concept in Keynesian economics is "money illusion"—the tendency to make decisions based on nominal quantities rather than converting those figures into "real" terms by adjusting for inflation. Money illusion seems to be pervasive in some domains. In one study (Baker, Gibbs, and Holmstrom 1994) of wage changes in a large financial firm, only 200 of more than 60,000 wage changes were nominal decreases, but 15% of employees suffered real wage cuts over a 10-year period, and, in many years, more than half of wage increases were real declines. It appears that employees don't seem to mind if their real wage falls as long as their nominal wage does not fall. Shafir, Diamond, and Tversky (1997 and in this volume) demonstrate the pervasiveness of money illusion experimentally (see also Fehr and Tyran 2001) and sketch ways to model it.

LABOR ECONOMICS

A central puzzle in macroeconomics is involuntary unemployment—why can some people not find work (beyond frictions of switching jobs, or a natural rate of unemployment)? A popular account of unemployment posits that wages are deliberately paid above the market-clearing level, which creates an excess supply of workers and hence, unemployment. But why are wages too high? One interpretation, "efficiency wage theory," is that paying workers more than they deserve is necessary to ensure that they have something to lose if they are fired, which motivates them to work hard and economizes on monitoring. Akerlof and Yellen (1990 and in this volume) have a different interpretation: Human instincts to reciprocate transform the employer-worker relation into a "gift-exchange." Employers pay more than they have to as a gift; and workers repay the gift by working harder than necessary. They show how gift-exchange can be an equilibrium (given reciprocal preferences), and show some of its macroeconomic implications.

In labor economics, gift-exchange is clearly evident in the elegant series of experimental labor markets described by Fehr and Gächter (2000 and in this volume). In their experiments, there is an excess supply of workers. Firms offer wages; workers who take the jobs then choose a level of effort, which is costly to the workers and valuable to the firms. To make the experiments interesting, firms and workers can enforce wages, but not effort levels. Since workers and firms are matched anonymously for just one period, and do not learn each other's identities, there is no way for either side to build reputations or for firms to punish workers who choose low effort. Self-interested workers should shirk, and firms should anticipate this and pay a low wage. In fact, firms deliberately pay high wages as gifts, and workers choose higher effort levels when they take higher-wage jobs. The strong correlation between wages and effort is stable over time.

Other chapters in this section explore different types of departures from the standard assumptions that are made about labor supply. For example, standard life-cycle theory assumes that, if people can borrow, they should prefer wage profiles that maximize the present value of lifetime wages. Holding total wage payments constant, and assuming a positive real rate of interest, present value maximization implies that workers should prefer declining wage profiles over increasing ones. In fact, most wage profiles are clearly rising over time, a phenomenon that Frank and Hutchens (1993 and in this volume) show cannot be explained by changes in marginal productivity. Rather, workers derive utility from positive changes in consumption but have self-control problems that would prevent them from saving for later consumption if wages were front-loaded in the life cycle. In addition, workers seem to derive positive utility from increasing wage profiles, per se, perhaps because rising wages are a source of self-esteem; the desire for increasing payments is much weaker for nonwage income (see Loewenstein and Sicherman 1991).

The standard life-cycle account of labor supply also implies that workers should intertemporally substitute labor and leisure based on the wage rate that they face and the value that they place on leisure at different points in time. If wage fluctuations are temporary, workers should work long hours when wages are high and short hours when wages are low. However, because changes in wages are often persisting, and because work hours are generally fixed in the short run, it is in practice typically difficult to tell whether workers are substituting intertemporally (though see Mulligan 1998). Camerer et al. (1997 and in this volume) studied labor supply of cab drivers in New York City. Cab drives represent a useful source of data for examining intertemporal substitution because drivers rent their cabs for a half-day and their work hours are flexible (they can quit early, and often do), and wages fluctuate daily because of changes in weather, day-of-the-week effects, and so forth. Their study was inspired by an alternative to the substitution hypothesis: Many drivers say that they set a daily income target and quit when they reach that target (in behavioral economics language, they isolate their daily decision and are averse to losing relative to an income target). Drivers who target daily will drive longer hours on low-wage days and quit early on high-wage days. This behavior is exactly the *opposite* of intertemporal substitution. Camerer et al. found that data from three samples of inexperienced drivers support the daily targeting prediction. But experienced drivers do not have negative elasiticies, either because target-minded drivers earn less and self-select out of the sample of experienced drivers, or drivers learn over time to substitute rather than target.

Perhaps the simplest prediction of labor economics is that the supply of labor should be upward sloping in response to a transitory increase in wage. Gneezy and Rustichini (this volume) document one situation in which this is not the case. They hired students to perform a boring task and either paid them a low piece-rate, a moderately high piece-rate, or no piece-rate at all. The surprising finding was that individuals in the low piece-rate condition produce the lowest "output" levels. Paying subjects, they argued, caused subjects to think of themselves as working in exchange for money and, when the amount of money was small, they decided that it simply wasn't worth it. In another study reported in their chapter, they showed a similar effect in a natural experiment that focused on a domain other than labor supply. To discourage parents from picking their children up late, a day-care center instituted a fine for each minute that parents arrived late at the center. The fine had the perverse effect of *increasing* parental lateness. The authors postulated that the fine eliminated the moral disapprobation associated with arriving late (robbing it of its gift-giving quality) and replaced it with a simple monetary cost that some parents decided was worth incurring. Their results show that the effect of price changes can be quite different than in economic theory when behavior has moral components that wages and prices alter.

FINANCE

In finance, standard equilibrium models of asset pricing assume that investors care about asset risks only if they affect marginal utility of consumption, and they incorporate publicly available information to forecast stock returns as accurately as possible (the "efficient markets hypothesis"). While these hypotheses do make some accurate predictions—e.g., the autocorrelation of price changes is close to zero—there are numerous anomalies. The anomalies have inspired the development of "behavioral finance" theories exploring the hypothesis that some investors in assets have limited rationality. Important articles are collected in Thaler (1993) and reviewed in Shleifer (2000), Hirshleifer (2001), and Barberis and Thaler in press.

An important anomaly in finance is the "equity premium puzzle": Average returns to stocks are much higher than returns to bonds (presumably to compensate stockholders for higher perceived risks).[23] To account for this pattern, Benartzi and Thaler (1995 and in this volume) assume a combination of decision isolation—investors evaluate returns using a 1-year horizon—and aversion to losses. These two ingredients create much more perceived risk to holding stocks than would be predicted by expected utility. Barberis, Huang, and Santos (2001) use a similar intuition in a standard asset-pricing equation. Several recent papers (Barberis, Shleifer, and Vishny 1998) show how empirical patterns of short-term underreaction to earnings surprises, and long-term overreaction, can arise from a quasi-Bayesian model.

Another anomaly is the magnitude of volume in the market. The so-called "Groucho Marx" theorem states that people should not want to trade with people who would want to trade with them, but the volume of stock market transactions is staggering. For example, Odean (1999 and in this volume) notes that the annual turnover rate of shares on the New York Stock Exchange is greater than 75%, and the daily trading volume of foreign-exchange transactions in all currencies (including forwards, swaps, and spot transactions) is equal to about one-quarter of the total annual world trade and investment flow. Odean then presents data on individual trading behavior which suggests that the extremely high volume may be driven, in part, by overconfidence on the part of investors.

The rise of behavioral finance is particularly striking because, until fairly recently, financial theory bet all its chips on the belief that investors are so rational that any observed historical patterns that can be used to beat the market are detected—the "efficient markets hypothesis." Early heretics like Shiller (1981), who argued empirically that stock-price swings are too volatile to reflect only news, and DeBondt and Thaler (1985), who discovered an important overreaction effect based on the psychology of representativeness, had their statistical work

[23] The idea of loss-aversion has appeared in other guises without being directly linked to its presence in individual choice. For example, Fama (1991, p. 1596) wrote that "consumers live in morbid fear of recessions." His conjecture can be reasonably construed only as a disproportionate aversion to a drop in standard of living, or overweighting the low probability of economic catastrophe. Both are features of prospect theory.

"audited" with special scrutiny (or worse, were simply ignored). In 1978 Jensen called the efficient markets hypothesis "the most well-established regularity in social science." Shortly after Jensen's grand pronouncement, the list of anomalies began to grow. (To be fair, anomaly-hunting is aided by the fact that market efficiency is such a precise, easily testable claim.) A younger generation are now eagerly sponging up as much psychology as they can to help explain, in a unified way, limits on the efficiency of markets.

LAW

A rapidly growing area of research is the application of behavioral economics to law (see Jolls, Sunstein, and Thaler 1998; Sunstein 2000). Legal decisions may be particularly influenced by limits on cognition because they are often made by individuals (e.g., judges) or groups (e.g., juries), without the influences of organizational aggregation or market discipline. In one of the earliest contributions, McCaffrey (1994) shows how cognitive framing by voters influences the structure of taxation. Guthrie, Rachlinski, and Wistrich (2001) find that judges exhibit biases in decision making (e.g., overconfidence about whether decisions will be overturned on appeal) similar to those of student subjects. Applying concepts from psychophysics, Kahneman, Schkade, and Sunstein (1998) show that hypothetical jurors' awards of punitive damages are very similar when expressed on a numerical six-point scale of outrage. But awards are highly variable when mapped to dollars, because there is no natural "modulus" for mapping outrage to money and different jurors use different mappings.

Applications of behavioral economics also thrive because the economic approach to law provides a useful source of benchmark predictions against which behavioral approaches can be contrasted. A good example is the Coase theorem. Coase noted that if two agents can bargain to efficiency, the assignment of property rights to one agent or another will not affect what outcome will occur after the bargaining (though it will affect which party pays or gets paid). From an efficiency perspective, this principle reduces pressure on the courts to "get it right." Whatever judgment the court arrives at, parties will quickly and efficiently negotiate to transfer property rights to the party that can make the best use of them. But if preferences are reference-dependent, and the legal assignment of property rights sets a reference point, then the Coase theorem is wrong: The unassigned party will often not pay as much as the property right-owner demands, even if the unassigned party would have done so *ex ante*, or would have benefited more from having been assigned the property right.

Jolls et al. note that behavioral concepts provide a way to address constructively concerns that laws or regulations are paternalistic. If people routinely make an unconscious error or one that they regret, then rules that inform them of errors or protect them from making them will help. This line of argument suggests a form of paternalism that is "conservative"—a regulation should be irresistible if it can help some irrational agents, and does little harm to rational ones (see Camerer et al. 2003). An example is "cooling-off" periods for high-pressure sales: People who are easily seduced into buying something they regret have a few days to renege on their agreement, and cool-headed rational agents are not harmed at all. Behavioral science can help inform what sorts of mistakes might be corrected this way.

New Foundations

In a final, brief section of the book, we include two papers that take behavioral economics in new directions. The first is case-based decision theory (Gilboa and Schmeidler 1995 and in this volume). Because of the powerful influence of decision theory (à la Ramsey, de Finetti, and Savage), economists are used to thinking of risky choices as inevitably reflecting a probability-weighted average of the utility of their possible consequences. The case-based approach starts from different primitives. It treats a choice situation as a "case" that has degrees of similarity to previous cases. Actions in the current case are evaluated by a sum or average of the outcomes of the same action in previous cases, weighted by the similarity of those previous cases to the current one. Cased-based theory substitutes the psychology of probability of future outcomes for a psychology of similarity with past cases.

The primitive process of case comparison is widely used in cognitive science and is probably a better representation of how choices are made in many domains than is probability-weighted utility evaluation. In hiring new faculty members or choosing graduate students, you probably don't talk in terms of utilities and probabilities. Instead, it is irresistible to compare a candidate to others who are similar and who did well or poorly. Case-based reasoning may be just as appealing in momentous decisions, like choosing a presidential ticket (Lloyd Bentsen's "I knew John Kennedy, and you're no John Kennedy") or managing international conflict ("Will fighting the drug war in Colombia lead to another Vietnam?"). Explicitly case-based approaches are also widely used in the economy. Agents base a list price for a house on the selling prices of nearby houses that are similar ("comparables"). "Nearest-neighbor" techniques based on similarity are also used in credit-scoring and other kinds of evaluations.

Another promising new direction is the study of emotion, which has boomed in recent years (see Loewenstein and Lerner 2001, for a review of this literature with a special focus on its implications for decision making). Damasio (1994) found that people with relatively minor emotional impairments have trouble making decisions and, when they do, they often make disastrous ones. Other research shows that what appears to be deliberative decision making may actually be driven by gut-level emotions or drives, then rationalized as a thoughtful decision (Wegner and Wheatley 1999). Loewenstein (1996 and in this volume, and 2000) discusses the possibilities and challenges from incorporating emotions into economic models. Behavioral economics is taking many other new directions that, we hope, will provide more than adequate content for a sequel to this volume in the not-too-distant future. One such thrust is the study of "hedonics" (e.g., Kahneman, Diener, and Schwartz 1999; Kahneman, Wakker and Sarin 1997). Hedonics begins by expanding the notion of utility. In the neoclassical view, utility is simply a number that codifies an expressed preference ("decision utility"). But people may also have memories of which goods or activities they enjoyed most ("remembered utility"), immediate momentary sensations of pleasure and pain ("instant utility"), and guesses about what future utilities will be like ("forecasted utility"). It would be remarkable coincidence if the human brain were built to guarantee that all four types of utility were exactly the same. For example, current utilities and decision processes both depend on emotional or visceral states (like hunger, fatigue, anger, sympathy, or arousal), and people overestimate the extent to which they will be in the same hedonic state in the future (Loewenstein 1996 and in this volume). As a result, forecasted utility is biased in the direction of instant utility (see Loewenstein, O'Donoghue, and Rabin 1999). The differences among these utilities is important because a deviation between decision utility and one of the other types of utility means that there is a mismatch which could perhaps be corrected by policies, education, or social guidance. For example, addicts may relapse because their remembered utility from using drugs highlights pleasure and excludes the instant disutility of withdrawal. The new hedonics links survey ratings of happiness with economic measures. For example, Easterlin (1974) stressed that average expressed ratings of happiness rise over decades much less than income rose. He suggested that people derive much of their happiness from relative income (which, by definition, cannot rise over time). Studies of worker quit rates, suicide, and other behavioral measures show similar effects of relative income and tie the happiness research to important economic phenomena (Clark and Oswald 1994, 1996; Frey and Stutzer 2002; Oswald 1997).

A third direction uses neuroscientific evidence to guide assumptions about economic behavior. Neuroscience is exploding with discoveries because of advances in imaging techniques that permit more precise temporal and spatial location of brain activity.[24] It is undoubtedly a large leap from precise neural activity to big decisions like planning for retirement or buying a car. Nonetheless, neuroscientific data may show that cognitive activities that are thought to be equivalent in economic theory actually differ, or that activities thought to be different may be the same. These data could resolve years or decades of debate that are difficult to resolve with other sorts of experiments (see Camerer, Loewenstein, and Prelec 2003).

A fourth direction acknowledges Herb Simon's emphasis on "procedural rationality" and models the procedures or algorithms that people use (e.g., Rubinstein 1998). This effort is likely to yield models that are not simply generalizations of standard ones. For example, Rubinstein (1988) models risky choice as a process

[24] A substantial debate is ongoing in cognitive psychology about whether knowing the precise details of how the brain carries out computations is necessary to understand functions and mechanisms at higher levels. (Knowing the mechanical details of how a car works may not be necessary to turn the key and drive it.) Most psychology experiments use indirect measures like response times, error rates, self-reports, and "natural experiments" due to brain lesions, and have been fairly successful in codifying what we know about thinking; pessimists think that brain scan studies won't add much. The optimists think that the new tools will inevitably lead to some discoveries and the upside potential is so great that they cannot be ignored. We share the latter view.

of comparing the similarity of the probabilities and outcomes in two gambles, and choosing on dimensions that are dissimilar. This procedure has some intuitive appeal but it violates all the standard axioms and is not easily expressed by generalizations of those axioms.

Conclusions

As we mentioned above, behavioral economics simply rekindles an interest in psychology that was put aside when economics was formalized in the latter part of the neoclassical revolution. In fact, we believe that many familiar economic distinctions do have a lot of behavioral content—they are *implicitly* behavioral and could surely benefit from more explicit ties to psychological ideas and data.

An example is the distinction between short-run and long-run price elasticity. Every textbook mentions this distinction, with a casual suggestion that the long run is the time it takes for markets to adjust, or for consumers to learn new prices, after a demand or supply shock. Adjustment costs undoubtedly have technical and social components, but they probably also have some behavioral underpinning in the form of gradual adaptation to loss as well as learning.

Another macroeconomic model that can be interpreted as implicitly behavioral is the Lucas "islands" model (1975). Lucas shows that business cycles can emerge if agents observe local price changes (on "their own island") but not general price inflation. Are the "islands" simply a metaphor for the limits of their own minds? If so, theory of cognition could add helpful detail (see Sims 2001).

Theories of organizational contracting are shot through with implicitly behavioral economics. Williamson (1985) and others motivate the incompleteness of contracts as a consequence of bounded rationality in foreseeing the future, but they do not tie the research directly to work on imagery, memory, and imagination. Agency theory begins with the presumption that there is some activity that the agent does not like to do—usually called "effort"—which cannot be easily monitored or enforced, and which the principal wants the agent to do. The term "effort" connotes lifting sides of beef or biting your tongue when restaurant customers are sassy. What exactly is the "effort" agents that dislike exerting and that principals want them to? It's not likely to be time on the job—if anything, workaholic CEOs may be working too hard! A more plausible explanation, rooted in loss-aversion, fairness, self-serving bias, and emotion, is that managers dislike making hard, painful decisions (such as large layoffs, or sacking senior managers who are close friends). Jensen (1993) hints at the idea that overcoming these behavioral obstacles is what takes "effort"; Holmstrom and Kaplan (2001) talk about why markets are better at making dramatic capital-allocation changes than managers and ascribe much of the managerial resistance to internal conflicts or "influence costs." Influence costs are the costs managers incur lobbying for projects that they like or personally benefit from (like promotions or raises). Influence costs are real but are also undoubtedly inflated by optimistic biases—each division manager really does think that his or her division desperately needs funds—self-serving biases, and social comparison of pay and benefits (otherwise, why are salaries kept so secret?).

In all these cases, conventional economic language has emerged that begs the deeper psychological questions of where adjustment costs, rigidities, mental "islands," contractual incompleteness, effort-aversion, and influence costs come from. Cognitively detailed models of these phenomena could surely produce surprising testable predictions.

Is Psychology Regularity an Assumption or a Conclusion?

Behavioral economics as described in this chapter, and compiled in this book, generally *begins* with assumptions rooted in psychological regularity and asks what follows from those assumptions. An alternative approach is to work backward, regarding a psychological regularity as a *conclusion* that must be proved, an *explanandum* that must be derived from deeper assumptions before we fully understand and accept it.

The alternative approach is exemplified by a fashionable new direction in economic theory (and psychology, too), which is to explain human behavior as the product of evolution (see *Journal of Economic Perspectives*, Spring 2002). Theories of this sort typically describe an evolutionary environment, a range of behaviors, and precise rules for evolution of behavior (e.g., replicator dynamics), and then show that a particular behavior is evolutionarily stable. For example, overconfidence about skill is evolutionarily adaptive under some conditions (Postlewaite and Comte 2001; Waldman 1994). Loss-aversion can be adaptive (because exaggerating one's preference for an object improves one's outcome under the Nash bargaining solution and perhaps other protocols; see Carmichael and MacLeod 1999). Rejections of low offers in take-it-or-leave-it ultimatum games are often interpreted as evidence of a specialized adaptation for punishing partners in repeated interactions, which cannot be "turned off" in unnatural one-shot games with strangers (Samuelson 2001).

We believe in evolution, of course, but we do not believe that behavior of intelligent, modern people immersed in socialization and cultural influence can be understood *only* by guessing what their ancestral lives were like and how their brains might have adapted genetically. The problem is that it is easy to figure out whether an evolutionary story identifies causes sufficient to bring about particular behavior, but it is almost impossible to know if those causes were the ones that actually did bring it about. So it is crucial, as with all models, to require the evolutionary stories to make falsifiable predictions and be consistent with as much available data as possible.[25] For example, the idea that rejections in one-shot ultimatum games come

[25] Winter and Zamir (1997) articulate the "unnatural habitat" viewpoint with remarkable precision. They write, "Although subjects fully understand the rules of the game and its payoff structure, their behavior is influenced by an unconscious perception that the situation they are facing is part of a much more extended game of similar real-life interactions." If the perception is truly unconscious, this account is immunized from falsification. For example, if subjects say, "I know the difference between a one-shot and a repeated game" (as most subjects do) their statements can be discounted if they are

from a repeated-game instinct that is genetically or culturally transmitted either predicts that behavior in one-shot and repeated ultimatum games will be the same or that players will learn to accept offers in one-shot games over time. The first prediction is clearly wrong and the second is only weakly observed (see Camerer 2003, chap. 2). The evolutionary adaptation hypothesis also does not gracefully account for the facts that young children accept low offers but learn to reject them as they grow older, and that adults in some simple societies (e.g., the Machiguenga in Peru) *do* make and accept low offers.

Another potential problem with evolutionary reasoning is that most studies posit a special brain mechanism to solve a particular adaptive problem, but ignore the effect of how that mechanism constrains solution of *other* adaptive problems. (This is nothing more than the general equilibrium critique of partial equilibrium modeling, applied to the brain.) For example, agents who cannot instinctively distinguish between one-shot and repeated games would presumably be handicapped in many other sorts of decisions that require distinguishing unique and repeated situations, or accurately forecasting horizons (such as life-cycle planning), unless they have a special problem making distinctions among types of games.

There are other, nonevolutionary, models that treat psychological regularity as a conclusion to be proved rather than an assumption to be used.[26] Such models usually begin with an observed regularity, and reverse-engineer circumstances under which it can be optimal. Models of this sort appeal to the sweet tooth that economists have for deriving behavior from "first principles" and rationalizing apparent irrationality. Theories of this sort are useful behavioral economics, but only if they are held to the same high standards of all good models (and of earlier behavioral models): Namely, can they parsimoniously explain a *range* of data with one simple mechanism? And what *fresh predictions* do they make?

Final Thoughts

Critics have pointed out that behavioral economics is not a unified theory but is instead a collection of tools or ideas. This is true. It is also true of neoclassical economics. A worker might rely on a "single" tool—say, a power drill—but also

---

assumed to be unaware that they really *don't* know the difference. Winter and Zamir then conclude, "We believe that it is practically impossible to create laboratory conditions that would cancel out this effect and induce subjects to act as if they were facing an anonymous one-shot [ultimatum game]." Then how can the unnatural habitat theory be falsified?

[26] For example, one recent model (Benabou and Tirole 1999) derives overconfidence from hyperbolic time discounting. Agents, at time 0, face a choice at time 1 between a task that requires an immediate exertion of effort and a payoff delayed till time 2, which depends on their level of some skill. Agents know that, due to hyperbolic time discounting, some tasks that are momentarily attractive at time 0 will become unattractive at time 1. Overconfidence arises because they persuade themselves that their skill level—i.e., the return from the task—will be greater than it actually will be so as to motivate themselves to do the task at time 1. There may, however, be far more plausible explanations for the same phenomenon, such as that people derive utility directly from self-esteem. Indeed the same authors later proposed precisely such a model (Benabou and Tirole 2000).

use a wide range of drill bits to do various jobs. Is this one tool or many? As Arrow (1986) pointed out, economic models do not derive much predictive power from the single tool of utility-maximization. Precision comes from the drill bits—such as time-additive separable utility in asset pricing, including a child's utility into a parent's utility function, to explain bequests; rationality of expectations for some applications and adaptive expectations for others; homothetic preferences for commodity bundles; price-taking in some markets and game-theoretic reasoning in others; and so forth. Sometimes these specifications are even contradictory—for example, pure self-interest is abandoned in models of bequests, but restored in models of life-cycle savings; and risk-aversion is typically assumed in equity markets and risk-preference in betting markets. Such contradictions are like the "contradiction" between a Phillips-head and a regular screwdriver: They are different tools for different jobs. The goal of behavioral economics is to develop better tools that, in some cases, can do both jobs at once.

Economists like to point out the natural division of labor between scientific disciplines: Psychologists should stick to individual minds, and economists to behavior in games, markets, and economies. But the division of labor is only efficient if there is effective coordination, and all too often economists fail to conduct intellectual trade with those who have a comparative advantage in understanding individual human behavior. All economics rests on *some* sort of implicit psychology. The only question is whether the implicit psychology in economics is good psychology or bad psychology. We think it is simply unwise, and inefficient, to do economics without paying *some* attention to good psychology.

We should finally stress that behavioral economics is not meant to be a separate approach in the long run. It is more like a school of thought or a style of modeling, which should lose special semantic status when it is widely taught and used. Our hope is that behavioral models will gradually replace simplified models based on stricter rationality, as the behavioral models prove to be tractable and useful in explaining anomalies and making surprising predictions. Then strict rationality assumptions now considered indispensable in economics will be seen as useful special cases (much as Cobb-Douglas production functions or expected value maximization are now)—namely, they help illustrate a point that is truly established only by more general, behaviorally grounded theory.

References

Akerlof, George A. 1991. "Procrastination and Obedience." *American Economic Review*, 81(2): 1–19.

Akerlof, George A., and Janet L. Yellen. 1990. "The Fair Wage-Effort Hypothesis and Unemployment." *Quarterly Journal of Economics*, 105(2): 255–83.

Allais, Maurice. 1953. "Le comportement de l'homme rationnel devant le risque, critique des postulats et axiomes de l'école américaine." *Econometrica*, 21: 503–46.

Angeletos, George-Marie, Andreas Repetto, Jeremy Tobacman, and Stephen Weinberg. 2001. "The Hyperbolic Buffer Stock Model: Calibration, Simulation, and Empirical Evaluation." *Journal of Economic Perspectives*, 15(3): 47–68.

Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. "Coherent arbitrariness: Stable Demand Curves without Stable Preferences." *Quarterly Journal of Economics*, 118: 73–105.

Ariely, Dan, and Klaus Wertenbroch. 2002. "Procrastination, Deadlines, and Performance: Self-Control by Precommitment," *Psychological Science*, 13 (May): 219–24.

Arrow, Kenneth J. 1986. "Rationality of Self and Others in an Economic System." In *Rational Choice: The Contrast between Economics and Psychology*, edited by R. M. Hogarth and M. W. Reder, 201–215. Chicago: University of Chicago Press.

Asch, David A., James P. Patton, and John C. Hershey. 1990. "Knowing for the Sake of Knowing." *Medical Decision Making*, 10: 47–57.

Babcock, Linda, and George Loewenstein. 1997. "Explaining Bargaining Impasse: The Role of Self-serving Biases." *Journal of Economic Perspectives*, 11: 109–26.

Baker, George, Mice Gibbs, and Bengt Holmstrom. 1994. "The Wage Policy of a Firm." *Quarterly Journal of Economics*, 109: 921–55.

Barberis, Nick, Ming Huang, and Tanos Santos. 2001. "Prospect Theory and Asset Prices." *Quarterly Journal of Economics*, 116(1): 1–53.

Barberis, Nick, Andrei Shleifer, and Robert Vishny. 1998. "A Model of Investor Sentiment." *Journal of Financial Economics*, 49(3): 307–43.

Barberis, Nick, and Richard Thaler. In press. "A Survey of Behavioral Finance. In *Handbook of the Economics of Finance*, edited by G. Constantinides, M. Harris, and R. Stulz. Available online at http://gsbwww.uchicago.edu/fac/nicholas.barberis/research/.

Benabou, Roland, and Jean Tirole. 1999. "Self Confidence: Intrapersonal Strategies." Princeton working paper.

———. 2000. "Self Confidence and Social Interactions." Princeton working paper.

Benartzi, Shlano, and Richard H. Thaler. 1995 and in this volume. "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics*, 110(1): 73–92.

Benzion, Uri, Amnon Rapoport, and Jacob Yagil. 1989. "Discount Rates Inferred from Decisions: An Experimental Study. *Management Science*, 35: 270–84.

Bewley, Truman F. 1998. "Why not cut pay?" *European Economic Review*, 42(3–5): 459–90.

Blinder, Alan S., Elie Canetti, and David E. Lebow, and Jeremy B. Rudd. 1998. *Asking about Prices: A New Approach to Understanding Price Stickiness*. New York: Russell Sage Foundation.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1): 166–93.

Camerer, Colin F. 1987. "Do Biases in Probability Judgment Matter in Markets? Experimental evidence." *American Economic Review*, 77: 981–97.

———. 1989a. "Does the Basketball Market Believe in the 'Hot Hand'?" *American Economic Review*, 79: 1257–61.

———. 1989b. "An Experimental Test of Several Generalized Utility Theories." *Journal of Risk and Uncertainty*, 2: 61–104.

———. 1992. "Recent Tests of Generalized Utility Theories." In *Utility Theories: Measurement and Applications*, edited by W. Edwards. Cambridge: Cambridge University Press.

———. 2000. "Prospect Theory in the Wild: Evidence from the Field." In *Choices, values, and frames*, edited by D. Kahneman and A. Tversky. Cambridge: Cambridge University Press.

———. 2003. *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton: Princeton University Press.

Camerer, Colin F., Linda Babcock, George Loewenstein, and Richard Thaler. 1997 and in this volume. “Labor Supply of New York City Cab Drivers: One Day at a Time.” *Quarterly Journal of Economics*, 111: 408–41.

Camerer, Colin F., and Teck H. Ho. 1999. “Experience Weighted Attraction Learning in Normal-Form Games.” *Econometrica*, 67: 827–74.

Camerer, Colin F., Teck H. Ho, and Kuan Chong. 2003. “A Cognitive Hierarchy Theory of One-Shot Games.” *American Economic Review Papers and Proceedings*, May.

———. 2002. “Functional EWA: A One-Parameter Theory of Learning in Games.” Caltech working paper. Also available online at http://www.hss.caltech.edu/~camerer/camerer.html.

Camerer, Colin F., and Robin M. Hogarth. 1999. “The Effects of Financial Incentives in Economics Experiments: A Review and Capital-Labor-Production Framework.” *Journal of Risk and Uncertainty*, 19: 7–42.

Camerer, Colin F., Sam Issacharoff, George Loewenstein, Ted O’Donoghue, and Mathew Rabin. 2003. “Regulation for conservatives: Behavioral economics and the case for asymmetric paternalism.” *University of Pennsylvania Law Review*, 151(3): 1211–54.

Camerer, Colin F., George Loewenstein, and Drazen Prelec. 2003. “Neural Economics: How Neuroscience Can Inform Economics.” Caltech working paper.

Camerer, Colin F., and Richard Thaler. 1995. “Anomalies: Dictators, Ultimatums, and Manners.” *Journal of Economic Perspectives*, 9: 209–19.

Camerer, Colin F., and Martin W. Weber. 1992. “Recent Developments in Modelling Preferences: Uncertainty and Ambiguity.” *Journal of Risk and Uncertainty*, 5: 325–70.

Carmichael, H. Lorne, and Bentley MacLeod. 1999. “Caring about Sunk Costs: A Behavioral Solution to Hold-up Problems with Small Stakes.” USC Olin working paper no. 99–19.

Carroll, Chris. 2000. “Requiem for the Representative Consumer? Aggregate Implications of Microeconomic Consumption Behavior.” *American Economic Review, Papers and Proceedings*, 90(2): 110–15.

Carson, Richard T., and John Horowitz. 1990. “Discounting Statistical Lives.” *Journal of Risk and Uncertainty*, 3: 403–13.

Charness, Gary, and Mathew Rabin. 2002. “Understanding Social Preferences with Simple Tests.” *Quarterly Journal of Economics*, 117: 817–69.

Clark, Andrew E., and Andrew J. Oswald. 1994. “Unhappiness and unemployment.” *Economic Journal* 104(424): 648–59.

Clark, Andrew E., and Andrew J. Oswald. 1996. “Satisfaction and Comparison Income.” *Journal of Public Economics*, 61(3): 359–81.

Cooper, David J., John H. Kagel, Wei Lo, and Qilo L. Gu. 1999. “Gaming against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers.” *American Economic Review*, 89(4): 781–804.

Costa Gomes, Miguel, Vincent Crawford, and Bruno Broseta. 2001. “Experimental Studies of Strategic Sophistication and Cognition in Normal-form Games.” *Econometrica*, 69: 1193–235.

Crawford, Vincent. 1997 and in this volume. “Theory and Experiment in the Analysis of Strategic Interaction.” In *Advances in economics and econometrics: Theory and applications*, edited by D. Kreps and K. Wallis. Cambridge: Cambridge University Press.

Damasio, Antonio R. 1994. *Descartes’ Error: Emotion, Reason, and the Human Brain*. New York: Putnam.

De Bondt, Werner F. M., and Richard Thaler. 1985. “Does the Stock Market Overreact?” *Journal of Finance*, 40(3): 793–805.

Degeorge, Francois, Jay Patel, and Richard Zeckhauser. 1999. “Earnings Management to Exceed Thresholds.” *Journal of Business*, 72: 1–33.

Easterlin, Richard A. 1974. “Does Economic Growth Improve the Human Lot? Some Empirical Evidence. In *Nations and households in economic growth: Essays in honor of Moses Abramowitz*, edited by P. A. David and M. W. Reder. New York: Academic Press.

Ellsberg, Daniel. 1961. “Risk, ambiguity, and the savage axioms.” *Quarterly Journal of Economics*, 75: 643–69.

Engelmann, Dirk, and Jean-Robert Tyran. 2002. “To Buy or Not to Buy? An Experimental Study of Consumer Boycotts in Retail Markets.” Discussion paper 2002–13, Department of Economics, University of St. Gallen. Available online at http://econpapers.hhs.se/paper/usgdp2002/2002-13.htm.

Epstein, Larry G., and Tan Wang. 1994. “Intertemporal Asset Pricing under Knightian Uncertainty.” *Econometrica*, 62(2): 282–322.

Fama, Eugene F. 1991. “Efficient Capital Markets: II.” *Journal of Finance*, 46(5): 1575–617.

Fehr, Ernst, and Simon Gächter. 2000 and in this volume. “Fairness and Retaliation: The Economics of Reciprocity.” *Journal of Economic Perspectives*, 14(3): 159–81.

Fehr, Ernst, and Klaus M. Schmidt. 1999. “A Theory of Fairness, Competition and Cooperation.” *Quarterly Journal of Economics*, 117(3): 817–68.

Fehr, Ernst, and Jean-Robert Tyran. 2001. “Does Money Illusion Matter?” *American Economic Review*, 91(5): 1239–62.

Frank, Robert H., and Robert M. Hutchens. 1993 and in this volume. “Wages, Seniority, and the Demand for Rising Consumption Profiles.” *Journal of Economic Behavior and Organizations*, 21: 251–76.

Frederick, Shane, George Loewenstein, and Ted O’Donoghue. 2002 and in this volume. “Intertemporal Choice: A Critical Review.” *Journal of Economic Literature*, 40(2): 351–401.

Frey, Bruno S., and Alois Stutzer. 2002. “What Can Economists Learn from Happiness Research.” *Journal of Economic Literature*, 40: 402–35.

Fudenberg, Drew, and David Levine. 1998. *The theory of Learning in Games*. Cambridge: MIT Press.

Ganguly, Ananda, John H. Kagel, and Donald Moser. 2000. “Do Asset Market Prices Reflect Traders’ Judgment Biases?” *Journal of Risk and Uncertainty*, 20: 219–46.

Geanakoplos, John, David Pearce, and Ennio Stacchetti. 1989. Psychological Games and Sequential Rationality.” *Games and Economic Behavior*, 1: 60–80.

Genesove, David, and Chris Mayer. 2001. “Loss Aversion and Seller Behavior: Evidence from the Housing Market.” *Quarterly Journal of Economics*, 116(4): 1233–60.

Ghirardato, Paolo, and Jonathan Katz. 2000. *Indecision Theory: Explaining Selective Abstention in Multiple Elections*. Caltech University working paper. Also available online at http://masada.hss.caltech.edu/~paolo/ghiro.html.

Gigerenzer, Gerd, Wolfang Hell, and Hartmut Blank. 1988. “Presentation and Content: The Use of Base Rates as a Continuous Variable.” *Journal of Experimental Psychology*, 14(3): 513–25.

Gilboa, Itzhak, and David Schmeidler, 1995. “Case-based Decision Theory.” *Quarterly Journal of Economics*, 110: 605–39.

Gilovich, Thomas, Dale W. Griffin, and Daniel Kahneman. 2002. *Heuristics of Judgment: Extensions and applications.* New York: Cambridge University Press.

Gilovich, Thomas, Robert Vallone, and Amos Tversky. 1985. “The Hot Hand in Basketball: On the Misperception of Random Sequences.” *Cognitive Psychology*, 17: 295–314.

Goeree, Jacob K., and Charles A. Holt. 1999. "Stochastic Game Theory: For Playing Games, not just for Doing Theory." *Proceedings of the National Academy of Sciences*, 96: 10564–67.

Goldstein, William M., and Ellad U. Weber. 1995. "Content and Its Discontents: The Use of Knowledge in Decision Making." In *Decision Making from a Cognitive Perspective: The Psychology of Learning and Motivation*, edited by J. R. Busemeyer, R. Hastie and D. L. Medin. Vol. 32. New York: Academic Press.

Grether, David M., and Charles Plott. 1979. "Economic Theory of Choice and the Preference Reversal Phenomenon." *American Economic Review*, 69(4): 623–38.

Grinblatt, Mark, and Matti Keloharju. 2001. "Distance, Language and Culture Bias: The Role of Investor Sophistication." *Journal of Finance*, 56(3): 1053–73.

Güth, Werner, Rolf Schmittberger, and Bemd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*, 3: 367–88.

Guthrie, Chris, Jeffrey Rachlinski, and Andrew Wistrich. 2001. "Inside the Judicial Mind. *Cornell Law Review*, 86: 777.

Hansen, Lars P., Thomas Sargent, and Thomas Tallarini Jr. 2001. Robust Permanent Income and Pricing." *Review of Economic Studies*, 66: 873–907.

Hardie, Bruce G.S., Eric J. Johnson, and Peter S. Fader. 1993. "Reference Dependence, Loss-Aversion, and Brand Choice." *Marketing Science*, 12: 378–94.

Harless, David, and Colin F. Camerer. 1994. "The Predictive Utility of Generalized Expected Utility Theories." *Econometrica*, 62: 1251–90.

Heath, Chip, Rick P. Larrick, and Joshua Klayman. 1998. "Cognitive Repairs: How Organizational Practices Can Compensate for Individual Shortcomings." *Review of Organizational Behavior*, 20: 1–38.

Heilbroner, Robert L. 1982. "The Socialization of the Individual in Adam Smith." *History of Political Economy*, 14(3): 427–39.

Helson, Harry. 1964. *Adaptation Level Theory: An Experimental and Systematic Approach to Behavior*. New York: Harper and Row.

Hertwig, Raleh, and Adreas Ortmann. (2001). "Experimental Practices in Economics: A Methodological Challenge for Psychologists." *Behavioral and Brain Sciences*, 24(3): 383–403.

Hey, John. 1997. "Experiments and the Economics of Individual Decision Making under Risk and Uncertainty." In *Advances in Economics and Econometrics: Theory and Applications*, edited by D. M. Kreps and K. F. Wallis, Vol. 1, Cambridge: Cambridge University Press.

Hirshleifer, David. 2001. "Investor Psychology and Asset Pricing." *Journal of Finance*, 56(4): 1533–98.

Ho, Teck, and J. Kuan Chong. 2003. "A Parsimonious Model of SKU Choice." *Journal of Marketing Research*, 40(3), (forthcoming).

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon L. Smith. 1994. "Preferences, Property Rights and Anonymity in Bargaining Games. *Games and Economic Behavior*, 7: 346–80.

Hogarth, Robin M, and Hillel J. Einhorn. 1992. "Order Effects in Belief Updating: The Belief-Adjustment Model." *Cognitive Psychology*, 24(1): 1–55.

Hogarth, Robin M., and Melvin W. Reder. 1987. *Rational Choice: The Contrast between Economics and Psychology*. Chicago: University of Chicago Press.

Holcomb, James H., and Paul S. Nelson. 1992. "Another Experimental Look at Individual Time Preference." *Rationality and Society*, 4(2): 199–220.

Holmstrom, Bengt, and Steve N. Kaplan. 2001. "Corporate Governance and Merger Activity in the U.S.: Making Sense of the 80s and 90s." *Journal of Economic Perspectives*, 15: 121–44.

Horowitz, John K. 1999. "A Test of Intertemporal Consistency." *Journal of Economic Behavior and Organization*, 17: 171–82.

Hsee, Chris K., and France Leclerc. 1998. "Will Products Look More Attractive when Evaluated Jointly or when Evaluated Separately?" *Journal of Consumer Research*, 25: 175–86.

Hsee, Chris K., George Loewenstein, Sally Blount, and Max Bazerman. 1999. "Preference Reversals between Joint and Separate Evaluations of Options: A Theoretical Analysis." *Psychological Bulletin*, 125(5): 576–90.

Huber, Joel, John W. Payne, and Chris Puto. 1982. "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *Journal of Consumer Research*, 9(1): 90–98.

Jensen, Michael. 1993. "The Modern Industrial Revolution, Exit, and the Failure of Internal Control System." *Journal of Finance*, 48(3): 831–81.

Johnson, Eric J., and David A. Schkade. 1989. "Bias in Utility Assessments: Further Evidence and Explanations." *Management Science*, 35: 406–24.

Jolls, Christine, Sunstein, Cass, and Richard Thaler. 1998. "A Behavioral Approach to Law and Economics." *Stanford Law Review*, 50: 1471–550.

Kahneman, Daniel, Ed Diener, and Norbert Schwartz, eds. 1999. "*Well-being: The Foundations of Hedonic Psychology*. New York: Russell Sage.

Kahneman, Daniel. J., and Shane Frederick. 2002. "Representativeness Revisited: Attribution Substitution in Intuitive Judgment." In *Heuristics of intuitive judgment: Extensions and applications*, edited by T. Gilovich, D. Griffin and D. Kahneman. New York: Cambridge University Press.

———. 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the market." *American Economic Review*, 76: 728–41.

———. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*, 98: 1325–48.

Kahneman, Daniel, David Schkade, and Cass Sunstein. 1998. "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages." *Journal of Risk and Uncertainty*, 16: 49–86.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47: 263–91.

Kahneman, Daniel, and Amos Tversky. 1982. "On the Study of Statistical Intuitions." In *Judgment under Uncertainty: Heuristics and Biases*, edited by D. Kahneman, P. Slovic, and A. Tversky. Cambridge: Cambridge University Press.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin. 1997. "Back to Bentham? Explorations of Experienced Utility." *Quarterly Journal of Economics*, 112: 375–405.

Knetsch, Jack L. 1992. "Preferences and Nonreversibility of Indifference Curves." *Journal of Economic Behavior and Organization*, 17(1): 131–39.

Koehler, Jay J. 1996. "Issues for the Next Generation of Base Rate Research." *Behavioral and Brain Sciences*, 19: 41–53.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112: 443–77.

Laibson, David, Andrea Repetto, and Jeremy Tobacman. 1998. "Self-control and Saving for Retirement." *Brookings Papers on Economic Activity*, 91–196.

Lewin, Sheri B. 1996. "Economics and Psychology: Lessons for Our Own Day from the Early Twentieth Century." *Journal of Economic Literature*, 34(3): 1293–1323.

List, John A., "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics* (February, 2003a), 118(1): 41–71.

Loewenstein, George. 1987. "Anticipation and the Valuation of Delayed Consumption." *Economic Journal*, 97: 666–84.

———. 1988. "Frames of Mind in Intertemporal Choice." *Management Science*, 34: 200–214.

———. 1992. "The Fall and Rise of Psychological Explanation in the Economics of Intertemporal Choice." In *Choice over time*, edited by G. Loewenstein and J. Elster. New York: Russell Sage.

———. 1996 and in this volume. "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes*, 65: 272–92.

———. 1999. "Experimental Economics from the Vantage-point of Behavioral Economics." *Economic Journal Controversy Corner: "What's the Use of Experimental Economics,"* 109: 25–34.

———. 2000. "Emotions in Economic Theory and Economic Behavior." *American Economic Review: Papers and Proceedings*, 90: 426–32.

———. 2001. "The Creative Destruction of Decision Research." *Journal of Consumer Research*, 28(3): 499–505.

Loewenstein, George, and Dan Adler. 1995. "A Bias in the Prediction of Tastes." *Economic Journal*, 105: 929–37.

Loewenstein, George, and Jon Elster. eds. 1992. *Choice over Time*. New York: Russell Sage.

Loewenstein, George, and Jennifer Lerner. 2003. "The Role of Emotion in Decision Making." In *The Handbook of Affective Science*, edited by R. J. Davidson, H. H, Goldsmith and K. R. Scherer. Oxford: Oxford University Press, 619–42.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 1999. *Projection Bias in the Prediction of Future Utility*. Working paper.

Loewenstein, George, and Drazen Prelec. 1992. "Anomalies in Intertemporal Choice: Evidence and an Interpretation." *Quarterly Journal of Economics* (May): 573–97.

———. 1993. "Preferences over Outcome Sequences." *Psychological Review*, 100(1): 91–108.

Loewenstein, George, Daniel Read, and Roy Baumeister. eds. 2003. *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*, Russell Sage Foundation, 13–86.

Loewenstein, George, and Nachum Sicherman. 1991. "Do Workers Prefer Increasing Wage Profiles?" *Journal of Labor Economics*, 9: 67–84.

Lucas, Robert E., Jr. 1975. "An Equilibrium Model of the Business Cycle." *Journal of Political Economy*, 83: 1133–44.

———. 1986. "Adaptive Behavior in Economic Theory." *Journal of Business*, 59(4): S401–S426.

MacKeigan, Linda D., L. N. Larson, Johaine R. Draugalis, J. Lyle Bootman, and L. R. Bruns. 1993. "Time Preference for Health Gains versus Health Losses." *PharmacoEconomics*, 3: 374–86.

Markowitz, Harry. 1952. "The Utility of Wealth." *Journal of Political Economy*, 60: 151–58.

McCabe, Kevin, Daniel Houser, Lee Ryan, Vernon Smith, and Theodore Trouard. 2001. "A Functional Imaging Study of Cooperation in Two-Person Reciprocal Exchange." *Proceedings of the National Academy of Sciences*, (98)20: 11832–35.

McCaffrey, Ed. 1994. "Cognitive Theory and Tax." *UCLA Law Review,* 41: 1861.

McKelvey, Richard D., and Thomas R. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior*, 7: 6–38.

———. 1998. "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics*, 1: 9–41.

Mukherji, Sujay. 1998. "Ambiguity Aversion and Incompleteness of Contractual Form." *American Economic Review*, 88(5): 1207–31.

Mulligan, Casey B. 1998. *Substitution over Time: Another Look at Life Cycle Labor Supply*. NBER Working Paper No.w6585.

O'Curry, Susan. 1999. "Consumer Budgeting and Mental Accounting." In *The Elgar Companion to Consumer Research and Economic Psychology*, edited by P. E. Earl and S. Kemp. Northhampton, MA: Cheltenham.

Odean, Terrance. 1998. "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance*, 53(5): 1775–98.

———. 1999 and in this volume. "Do Investors Trade Too Much?" *American Economic Review*, 89: 1279–98.

O'Donoghue, Ted, and Mathew Rabin. 1999 and in this volume. "Doing It Now or Later." *American Economic Review*, 89: 103–24.

———. 2001. "Choice and Procrastination." *Quarterly Journal of Economics*, 116(1): 121–60.

Oswald, Andrew. 1997. "Happiness and Economic Performance." *Economic Journal*, 107: 1815–31.

Payne, John W., Jane R. Bettman, and Eric J. Johnson. 1992. "Behavioral Decision Research: A Constructive Processing Perspective." *Annual Review of Psychology*, 43: 87–131.

Pender, John L. 1996. "Discount Rates and Credit Markets: Theory and Evidence from Rural India." *Journal of Development Economics*, 50(2): 257–96.

Phelps, Edmund S., and Robert A. Pollak. 1968. "On Second-Best National Saving and Game-Equilibrium Growth." *Review of Economic Studies*, 35: 185–99.

Postlewaite, Andrew, and Oliver Compte. 2001. "Confidence Enhanced Performance." Working Paper, University of Pennsylvania. Also at http://www.ssc.upenn.edu/%7Eapostlew/paper/working.html.

Prelec, Drazen, and George Loewenstein. 1991. "Decision Making over Time and under Uncertainty: A common approach." *Management Science*, 37: 770–86.

———. 1998. "The Red and the Black: Mental Accounting of Savings and Debt." *Marketing Science*, 17: 4–28.

Prelec, Drazen, Birger Wernerfelt, and Florian Zettelmeyer. 1997. "The Role of Inference in Context Effects: Inferring What You Want from What Is Available." *Journal of Consumer Research*, 24(1): 118–25.

Quiggin, John. 1993. *Generalized Expected Utility Theory: The Rank-dependent Expected Utility Model.* Amsterdam: Kluwer-Nijhoff.

Rabin, Mathews. 1993 and in this volume. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5): 1281–302.

———. 2000. "Risk-Aversion and Expected Utility Theory: A Calibration Theorem." *Econometrica*, 68(5): 1281–92.

———. 2002. "The Law of Small Numbers." *Quarterly Journal of Economics*, 117: 775–816.

Rabin, Mathews, and Joel Schrag. 1999. "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics*, 114(1): 37–82.

Ramsey, Frank P. 1928. "A Mathematical Theory of Saving. "*Economic Journal*, 38(152): 543–59.

Read, Daniel, George Loewenstein, and Mathew Rabin. 1999. "Choice Bracketing." *Journal of Risk and Uncertainty*, 19: 171–97.

Rose, Naney, and Catherine Wolfram. 2000. "Has the "Million-Dollar Cap" Affected the CEO Pay?" *American Economic Review*, 90(2): 197–202.

Roth, Airin, Ide Erev, Robert Slonim, and Greg Barron. 2000. *Equilibrium and Learning in Economic Environments: The Predictive Value of Approximations*. Discussion paper, Harvard University, Department of Economics.

Rubinstein, Ariel. 1988. "Similarity and Decision-making under Risk (Is There a Utility Theory Resolution to the Allais Paradox?)." *Journal of Economic Theory*, 46: 145–53.

Samuelson, Larry. 2001. "Adaptations, Analogies and Anomalies." *Journal of Economic Theory*, 97: 320–66.

Samuelson, Paul. 1937. "A Note on Measurement of Utility." *Review of Economic Studies*, 4: 155–61.

Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: Wiley.

Sayman, Serdar A. and Ayse Öncüler, 1997. "A Meta Analysis of the Willingness-to-Accept and Willingness-to-Pay Disparity." Working paper, Wharton School, University of Pennsylvania. Available at http://home.ku.edu.tr/~ssayman/research/meta.pdf.

Schelling, Thomas C. 1978. "Economics, or the Art of Self-management." *American Economic Review*, 68(2): 290–94.

Schlag, Karl. 1998. "Why Imitate, and If So, How? A Bounded Rational Approach to Multi-armed Bandits." *Journal of Economic Theory*, 78(1): 130–56.

Shafir, Eldar, Peter Diamond, and Amos Tversky, 1997, and in this volume. "Money Illusion." *Quarterly Journal of Economics*, 112(2): 341–74.

Shefrin, Hersh M., and Richard H. Thaler. 1992. "Mental Accounting, Saving, and Self-control." In *Choice over Time*, G. Loewenstein and J. Elster. New York: Russell Sage.

Shiller, Robert. 1981. "Do Stock Prices Move too Much to Be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71: 421–36.

Shleifer, Andres. 2000. *Inefficient Markets*. Oxford: Oxford University Press.

Simonson, Itamer, 1990. "The Effect of Purchase Quantity and Timing on Variety-Seeking Behavior." *Journal of Marketing Research*, 17: 150–62.

Simonson, Itamer, and Amos Tversky. 1992. "Choice in Context: Trade-off Contrast and Extremeness Aversion." *Journal of Marketing Research*, 29: 281–95.

Simonson, Itamer, and Russell S. Winer, 1992. "The Influence of Purchase Quantity and Display Format on Consumer Preference for Variety." *Journal of Consumer Research*, 19: 133–38.

Sims, Christopher. 2001. "Implications of Rational Inattention." Unpublished paper, Princeton University.

Slovic, Paul. 1995. "The Construction of Preferences." *American Psychologist*, 50: 364–71.

Smith, Adam. 1759/1892. "The Theory of Moral Sentiments." New York: Prometheus.

Stahl, Dale, and Paul Wilson. 1995. "On Players' Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior*, 10(1): 218–54.

Starmer, Chris. 2000. "Developments in Non-expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk." *Journal of Economic Literature*, 38: 332–82.

Stigler George. 1965. "The Development of Utility Theory," in *Essays in the History of Economics*. Chicago: University of Chicago Press.

Strahilevitz, Michal, and George Loewenstein. 1998. "The Effects of Ownership History on the Valuation of Objects." *Journal of Consumer Research*, 25: 276–89.

Strotz, Robert H. 1955. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies*, 23: 165–80.

Sunstein, Cass. ed. 2000. *Behavioral Law and Economics*. Cambridge: Cambridge University Press.

Thaler, Richard. 1980. "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization*, 1: 39–60.

———. 1981. "Some Empirical Evidence on Dynamic Inconsistency." *Economics Letters*, 8: 201–7.

———. ed. 1993. *Advances in Behavioral Finance*. New York: Russell Sage.

———. 1994. "Psychology and Savings Policies." *American Economic Review*, 84(2): 186–92.

——— 1999 and in this volume. "Mental Accounting Matters." *Journal of Behavioral Decision Making*, 12: 183–206.

Thaler, Richard H., and Hersh M. Shefrin. 1981. "An Economic Theory of Self-control." *Journal of Political Economy*, 89(2): 392–406.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science*, 185: 1124–31.

———. 1981. "The Framing of Decisions and the Psychology of Choice." *Science*, 211: 453–58.

———. 1983. "Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review*, 90: 293–315.

———. 1991. "Loss aversion in Riskless Choice: A Reference-dependent Model." *Quarterly Journal of Economics*, 106: 1039–61.

———. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5: 297–323.

Tversky, Amos, Paul Slovic, and Daniel Kahneman. 1990. "The Causes of Preference Reversal." *American Economic Review*, 80: 204–17.

Van Boven, Leaf, David Dunning, and George Loewenstein. 2000. "Egocentric Empathy Gaps between Owners and Buyers: Misperceptions of the Endowment effect." *Journal of Personality and Social Psychology*, 79: 66–76.

Van Boven, Leaf, George Loewenstein, and David Dunning. 2000. "Biased Predictions of Others' Tastes: Underestimation of Owners' Selling Prices by 'buyer's agents.'" Working paper.

Waldman, Michael. 1994. "Systematic Errors and the Theory of Natural Selection." *The American Economic Review*, 84(3): 482–97.

Weisbuch, Gerard, Alan Kirman, and Dorathea Herreiner. 2000. "Market Organisation and Trading Relationships." *The Economic Journal*, 110: 411–36.

Weizsacker, George. In press. "Ignoring the Rationality of Others: Evidence from Experimental Normal-Form Games." *Games and Economic Behavior*.

Wegner, Dave M., and Thalia Wheatley. 1999. "Apparent Mental Causation: Sources of the Experience of Will." *American Psychologist*, 54(7): 480–92.

Williamson, Oliver. 1985. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.

Winter, Eyal, and Shmuel Zamir. 1997. "An Experiment with Ultimatum Bargaining in a Changing Environment." Hebrew University Center for the Study of Rationality working paper #159. Also at http://www.ratio.huji.ac.il/dp.asp.

PART II

Basic Topics

Reference-Dependence and Loss-Aversion

Preference Over Risky and Uncertain Outcomes

Intertemporal Choice

Fairness and Social Preferences

Game Theory

C H A P T E R 2

Experimental Tests of the Endowment Effect and the Coase Theorem

DANIEL KAHNEMAN, JACK L. KNETSCH,

AND RICHARD H. THALER

1. Introduction

The standard assumptions of economic theory imply that when income effects are small, differences between an individual's maximum willingness to pay (WTP) for a good and minimum compensation demanded for the same entitlement (willingness to accept [WTA]) should be negligible (Willig 1976). Thus indifference curves are drawn without reference to current endowments; any difference between equivalent and compensating variation assessments of welfare changes is in practice ignored;[1] and there is wide acceptance of the Coase theorem assertion that, subject to income effects, the allocation of resources will be independent of the assignment of property rights when costless trades are possible.

The assumption that entitlements do not affect value contrasts sharply with empirical observations of significantly higher selling than buying prices. For example, Thaler (1980) found that the minimal compensation demanded for accepting a .001 risk of sudden death was higher by one or two orders of magnitude than the amount people were willing to pay to eliminate an identical existing risk. Other examples of similar reported findings are summarized in table 2.1. The disparities observed in these examples are clearly too large to be explained plausibly by income effects.

Several factors probably contribute to the discrepancies between the evaluations of buyers and sellers that are documented in table 2.1. The perceived illegitimacy of the transaction may, for example, contribute to the extraordinarily high demand for personal compensation for agreeing to the loss of a public good (e.g., Rowe, d'Arge, and Brookshire 1980). Standard bargaining habits may also contribute to a discrepancy between the stated reservation prices of buyers and

Financial support was provided by Fisheries and Oceans Canada, the Ontario Ministry of the Environment, and the behavioral economics program of the Alfred P. Sloan Foundation. We wish to thank Vernon Smith for encouraging us to conduct these experiments and for providing extensive comments on earlier drafts. Of course, the usual disclaimer applies.

[1] For example, the conventional prescription for assessing environmental and other losses is that, "practically speaking, it does not appear to make much difference which definition is accepted" (Freeman 1979, p. 3).

TABLE 2.1
Summary of Past Tests of Evaluation Disparity

	Means			*Medians*		
Study and Entitlement	*WTP*	*WTP*	*Ratio*	*WTP*	*WTA*	*Ratio*
Hypothetical surveys:						
Hammack and Brown (1974): marshes	$247	$1,044	4.2			
Sinclair (1978): fishing				35	100	2.9
Banford et al. (1979):						
Fishing pier	43	120	2.8	47	129	2.7
Postal service	22	93	4.2	22	106	4.8
Bishop and Heberlein (1979): goose hunting permits	21	101	4.8			
Rowe et al. (1980): visibility	1.33	3.49	2.6			
Brookshire et al. (1980): elk hunting[a]	54	143	2.6			
Heberlein and Bishop (1985): deer hunting	31	513	16.5			
Real exchange experiments:						
Knetsch and Sinden (1984): lottery tickets	1.28	5.18	4.0			
Heberlein and Bishop (1985): deer hunting	25	172	6.9			
Coursey et al. (1987): taste of sucrose octa-acetate[b]	3.45	4.71	1.4	1.33	3.49	2.6
Brookshire and Coursey (1987): park trees[c]	10.12	56.60	5.6	6.30	12.96	2.1

[a] Middle-level change of several used in study.
[b] Final values after multiple iterations.
[c] Average of two levels of tree plantings.

sellers. Sellers are often rewarded for overstating their true value, and buyers for understating theirs (Knez, Smith, and Williams 1985). By force of habit they may misrepresent their true valuations even when such misrepresentation confers no advantage, as in answering hypothetical questions or one-shot or single transactions. In such situations the buying-selling discrepancy is simply a strategic mistake, which experienced traders will learn to avoid (Coursey, Hovis, and Schulze 1987; Brookshire and Coursey 1987).

The hypothesis of interest here is that many discrepancies between WTA and WTP, far from being a mistake, reflect a genuine effect of reference positions on preferences. Thaler (1980) labeled the increased value of a good to an individual when the good becomes part of the individual's endowment the "endowment effect." This effect is a manifestation of "loss aversion," the generalization that losses are weighted substantially more than objectively commensurate gains in the evaluation of prospects and trades (Kahneman and Tversky 1979; Tversky and Kahneman, in press). An implication of this asymmetry is that if a good is evaluated as a loss when it is given up and as a gain when it is acquired, loss aversion will, on average, induce a higher dollar value for owners than for potential buyers, reducing the set of mutually acceptable trades.

There are some cases in which no endowment effect would be expected, such as when goods are purchased for resale rather than for utilization. A particularly clear case of a good held exclusively for resale is the notional token typically traded in experimental markets commonly used to test the efficiency of market institutions (Plott 1982; Smith 1982). Such experiments employ the induced-value technique in which the objects of trade are tokens to which private redemption values that vary among individual participants have been assigned by the experimenter (Smith 1976). Subjects can obtain the prescribed value assigned for the tokens when redeeming them at the end of the trading period; the tokens are otherwise worthless.

No endowment effect would be expected for such tokens, which are valued only because they can be redeemed for cash. Thus both buyers and sellers should value tokens at the induced value they have been assigned. Markets for induced-value tokens can therefore be used as a control condition to determine whether differences between the values of buyers and sellers in other markets could be attributable to transaction costs, misunderstandings, or habitual strategies of bargaining. Any discrepancy between the buying and selling values can be isolated in an experiment by comparing the outcomes of markets for real goods with those of otherwise identical markets for induced-value tokens. If no differences in values are observed for the induced-value tokens, then economic theory predicts that no differences between buying and selling values will be observed for consumption goods evaluated and traded under the same conditions.

The results from a series of experiments involving real exchanges of tokens and of various consumption goods are reported in this paper. In each case, a random allocation design was used to test for the presence of an endowment effect. Half of the subjects were endowed with a good and became potential sellers in each market; the other half of the subjects were potential buyers. Conventional economic analysis yields the simple prediction that one-half of the goods should be traded in voluntary exchanges. If value is unaffected by ownership, then the distribution of values in the two groups should be the same except for sampling variation. The supply and demand curves should therefore be mirror images of each other, intersecting at their common median. The null hypothesis is, therefore, that half of the goods provided should change hands. Label this predicted volume V^*. If there is an endowment effect, the value of the good will be higher for sellers than for buyers, and observed volume V will be less than V^*. The ratio V/V^* provides a unit-free measure of the undertrading that is produced by the effect of ownership on value. To test the hypothesis that market experience eliminates undertrading, the markets were repeated several times.

A test for the possibility that observed undertrading was due to transaction costs was provided by a comparison of the results from a series of induced-value markets with those from the subsequent goods markets carried out with identical trading rules. Notice that this comparison can also be used to eliminate numerous other possible explanations of the observed undertrading. For example, if the instructions to the subjects are confusing or misleading, the effects should show up in both the induced-value markets and the experimental markets for real goods. Section 2 describes studies of trading volume in induced-value markets and in consumption goods markets. Section 3 provides a further test for strategic behavior and demonstrates that the disparity findings are not likely caused by this. Section 4 investigates the extent to which the undertrading of goods is produced by reluctance to buy and reluctance to sell. Section 5 examines undertrading in bilateral negotiations and provides a test of the Coase theorem. Section 6 describes an experiment that rules out income effects and a trophy effect as explanations of the observed valuation disparity. Implications of the observed effects are discussed in section 7.

2. Repeated Market Experiments

In experiment 1, 44 students in an advanced undergraduate law and economics class at Cornell University received a packet of general instructions plus 11 forms, one for each of the markets that were conducted in the experiment. (The instructions for all experiments are available from the authors.) The first three markets were conducted for induced-value tokens. Sellers received the following instructions (with differences for buyers in brackets):

> In this market the objects being traded are tokens. You are an owner, so you now own a token [You are a buyer, so you have an opportunity to buy a token] which has a value to you of $x. It has this value to you because the experimenter will give you this much money for it. The value of the token is different for different individuals. A price for the tokens will be determined later. For each of the prices listed below, please indicate whether you prefer to: (1) Sell your token at this price and receive the market price. [Buy a token at this price and cash it in for the sum of money indicated above.] (2) Keep your token and cash it in for the sum of money indicated above. [Not buy a token at this price.] For each price indicate your decision by marking an X in the appropriate column.

Part of the response form for sellers follows:

> At a price of $8.75 I will sell _____ I will not sell _____
>
> At a price of $8.25 I will sell _____ I will not sell _____

The same rectangular distribution of values—ranging from $0.25 to $8.75 in steps of $0.50—was prepared for both buyers and sellers. Because not all the forms were actually distributed, however, the induced supply and demand curves were not always precisely symmetrical. Subjects alternated between the buyer and seller role in the three successive markets and were assigned a different individual redemption value in each trial.

Experimenters collected the forms from all participants after each market period and immediately calculated and announced the market-clearing price,[2] the number of trades, and the presence or absence of excess demand or supply at the market-clearing price.[3] Three buyers and three sellers were selected at random after each of the induced markets and were paid off according to the preferences stated on their forms and the market-clearing price for that period.

Immediately after the three induced-value markets, subjects on alternating seats were given Cornell coffee mugs, which sell for $6.00 each at the bookstore. The experimenter asked all participants to examine a mug, either their own or their neighbor's. The experimenter then informed the subjects that four markets for mugs would be conducted using the same procedures as the prior induced markets with two exceptions: (1) One of the four market trials would subsequently be selected at random, and only the trades made on this trial would be executed. (2) In the binding market trial, *all* trades would be implemented, unlike the subset implemented in the induced-value markets.[4] The initial assignment of buyer and seller roles was maintained for all four trading periods. The clearing price and the number of trades were announced after each period. The market that "counted" was indicated after the fourth period, and transactions were executed immediately. All sellers who had indicated that they would give up their mugs for a sum at the market-clearing price exchanged their mugs for cash, and successful buyers paid this same price and received their mugs. This design was used to permit learning to take place over successive trials and yet make each trial potentially binding. The same procedure was then followed for four more successive markets using boxed ballpoint pens with a visible bookstore price tag of $3.98, which were distributed to the subjects who had been buyers in the mug markets.

For each goods market, subjects completed a form similar to that used for the induced-value tokens, with the following instructions:

> You now own the object in your possession. [You do not own the object that you see in the possession of some of your neighbors.] You have the option of selling it [buying

[2] The instructions stated that "*it is in your best interest to answer these questions truthfully*. For any question, treat the price as fixed. (In economics jargon, you should act as 'price takers'.)" All the subjects were junior and senior economics majors, and so they were familiar with the terms used. If subjects asked how the market prices were determined, they were told, truthfully, that the market price was the point at which the elicited supply and demand curves intersected. The uniformity of the results across many different experiments suggests that this information had no discernible effect on behavior. Furthermore, the responses of the subjects in the induced-value portion of the experiments indicate that nearly all understood and accepted their role as price takers. See also experiment 5, in which a random price procedure was used.

[3] When this occurred, a random draw determined which buyers and sellers were accommodated.

[4] The experimental design was intended to give the markets for consumption goods every possible chance to be efficient. While in the induced-value markets not everyone was paid, in the consumption goods markets everyone was paid. Also, the consumption goods markets were conducted after the induced-value markets and were repeated four times each, to allow the subjects the maximum opportunity for learning.

> one] if a price, which will be determined later, is acceptable to you. For each of the possible prices below indicate whether you wish to: (1) sell your object and receive this price [Pay this price and receive an object to take home with you], or (2) keep your object and take it home with you. [Not buy an object at this price.] For each price indicate your decision by marking an X in the appropriate column.

The buyers and sellers in the consumption goods markets faced the same incentives that they had experienced in the induced-value markets. Buyers maximized their potential gain by agreeing to buy at all prices below the value they ascribed to the good, and sellers maximized their welfare by agreeing to sell at all prices above the good's worth to them. As in the induced-value markets, it was in the best interest of the participants to act as price takers.

As shown in table 2.2, the markets for induced-value tokens and consumption goods yielded sharply different results. In the induced-value markets, as expected, the median buying and selling prices were identical. The ratio of actual to predicted volume (V/V^*) was 1.0, aggregating over the three periods. In contrast, the median selling prices in the mug and pen markets were more than twice the median buying prices, and the V/V^* ratio was only .20 for mugs and .41 for pens. Observed volume did not increase over successive periods in either the mug or the

TABLE 2.2
Results of Experiment 1

Induced-Value Markets				
Trial	*Actual Trades*	*Expected Trades*	*Price*	*Expected Price*
1	12	11	3.75	3.75
2	11	11	4.75	4.75
3	10	11	4.25	4.25
Consumption Goods Markets				
Trial	*Trades*	*Price*	*Median Buyer Reservation Price*	*Median Seller Reservation Price*
	Mugs (Expected Trades = 11)			
4	4	4.25	2.75	5.25
5	1	4.75	2.25	5.25
6	2	4.50	2.25	5.25
7	2	4.25	2.25	5.25
	Pens (Expected Trades = 11)			
8	4	1.25	.75	2.50
9	5	1.25	.75	1.75
10	4	1.25	.75	2.25
11	5	1.25	.75	1.75

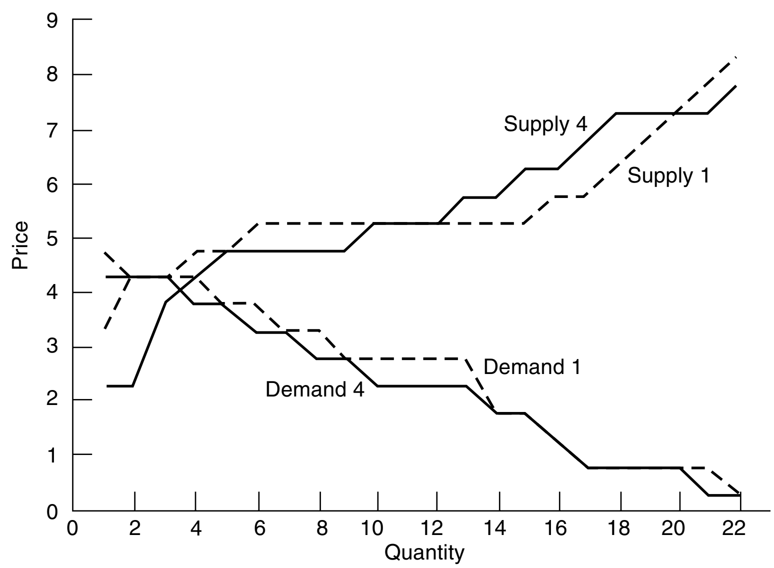


FIGURE 2.1 Supply and Demand curves, markets 1 and 4.

pen markets, providing no indication that subjects learned to adopt equal buying and selling prices.

The results of the first and last markets for coffee mugs are also displayed in figure 2.1. There are five features to notice in this figure: (1) Both buyers and sellers display a wide range of values, indicating that in the absence of an endowment effect there would be enough rents to produce gains from trade. Indeed, the range of values is similar to that used in the induced-value markets, which had near-perfect market efficiency. (2) The distribution of selling prices has a single mode, unlike some recent results in which an evaluation discrepancy could be explained by a bimodal distribution of compensation demanded (Boyce et al. 1990). (3) The payment of a small commission for trading, such as \$0.25 per trade, would not significantly alter the results. (4) The mugs were desirable. Every subject assigned a positive value to the mug, and the lowest value assigned by a seller was \$2.25. (5) Neither demand nor supply changed much between the first and last markets.

Experiment 2 was conducted in an undergraduate microeconomics class at Cornell (N = 38). The procedure was identical to that of experiment 1, except that the second consumption good was a pair of folding binoculars in a cardboard frame, available at the bookstore for \$4.00. The results are reported in table 2.3.

In experiments 3 and 4, conducted in Simon Fraser University undergraduate economics classes, the subjects were asked to provide minimum selling prices or maximum buying prices rather than to answer the series of yes or no questions used in experiments 1 and 2. The induced-value markets were conducted with no monetary payoffs and were followed by four markets for pens in experiment 3

TABLE 2.3
Results of Experiment 2

Induced-Value Markets				
Trial	*Actual Trades*	*Expected Trades*	*Price*	*Expected Price*
1	10	10	3.75	4.75
2	9	10	4.75	4.25
3	7	8	4.25	4.75
Consumption Goods Markets				
Trial	*Trades*	*Price*	*Median Buyer Reservation Price*	*Median Seller Reservation Price*
	Mugs (Expected Trades = 9.5)			
4	3	3.75	1.75	4.75
5	3	3.25	2.25	4.75
6	2	3.25	2.25	4.75
7	2	3.25	2.25	4.25
	Binoculars (Expected Trades = 9.5)			
8	4	1.25	.75	1.25
9	4	.75	.75	1.25
10	3	.75	.75	1.75
11	3	.75	.75	1.75

and five markets for mugs in experiment 4. In experiment 3, subjects were told that the first three markets for pens would be used for practice, so only the fourth and final market would be binding. In experiment 4, one of the five markets was selected at random to count, as in experiments 1 and 2. Other procedures were unchanged. The results are shown in table 2.4.

Experiments 2–4 all yielded results similar to those obtained in experiment 1. Summing over the induced-value markets in all four experiments produced a V/V^* index of .91. This excellent performance was achieved even though the participants did not have the benefit of experience with the trading rules, there were limited monetary incentives in experiments 1 and 2, and there were no monetary incentives in experiments 3 and 4. In the markets for consumption goods, in which all participants faced monetary incentives and experience with the market rules gained from the induced-value markets, V/V^* averaged .31, and median selling prices were more than double the corresponding buying prices. Trading procedures were precisely identical in markets for goods and for induced-value tokens. The high volume of trade in money tokens therefore eliminates transaction costs (or any other feature that was present in both types of markets) as an explanation of the observed undertrading of consumption goods.

TABLE 2.4
Results of Experiments 3 and 4

Trial	*N*	*Object*	*Actual Trades*	*Expected Trades*	*Ratio of Seller Median Value to Buyer Median Value*
			Experiment 3		
1	26	Induced	5	6.5	
2	26	Pen	2	6.5	6.0
3	26	Pen	2	6.5	6.0
4	26	Pen	2	6.5	5.0
5	26	Pen	1	6.5	5.0
			Experiment 4		
1	74	Induced	15	18.5	
2	74	Induced	16	18.5	
3	74	Mug	6	18.5	3.8
4	74	Mug	4	18.5	2.8
5	72	Mug	4	18	2.2
6	73	Mug	8	18	1.8
7	74	Mug	8	18.5	1.8

It should be noted that subjects in the position of buyers were not given money to use for purchases, but rather had to make transactions using their own money. (Subjects were told to bring money to class and that credit and change would be available if necessary. Some subjects borrowed from friends to make payments.) The aim was to study transactions in a realistic setting. While the present design makes potential sellers slightly wealthier, at least in the first market, the magnitude of the possible income effect is trivial. In one of the markets the equilibrium price was only \$0.75, and the prices in other markets were never above a few dollars. Also, as shown in experiments 7 and 8 below, equal undertrading was found in designs that eliminated the possibility of an income effect or cash constraint.

As shown in tables 2.1–2.4, subjects showed almost no undertrading even in their first trial in an induced-value market. Evidently neither bargaining habits nor any transaction costs impede trading in money tokens. On the other hand, there is no indication that participants in the markets for goods learned to make valuations independent of their entitlements. The discrepant evaluations of buyers and sellers remained stable over four, and in one case five, successive markets for the same good and did not change systematically over repeated markets for successive goods.

A difference in procedure probably explains the apparent conflict between these results and the conclusion reached in some other studies, that the WTA-WTP discrepancy is greatly reduced by market experience. The studies that reported a disciplinary effect of market experience assessed this effect by comparing the responses of buyers and sellers in preliminary hypothetical questions or nonbinding market trials to their behavior in a subsequent binding trial with real monetary payoffs (Knez et al. 1985; Brookshire and Coursey 1987; Coursey et al. 1987). In the present experiments, the markets for consumption goods were real and potentially binding from the first trial, and the WTA-WTP discrepancy was found to be stable over a series of such binding trials.

It should be stressed that previous research did not actually demonstrate that the discrepancy between buyers and sellers is eliminated in markets. Although the discrepancy between the final selling and buying prices in the sucrose octa-acetate experiment of Coursey et al. (1987) was not statistically significant, the ratio of median prices of sellers and buyers was still 2.6.[5] If the buyers and sellers had been allowed to trade according to their final bids, a total of nine advantageous exchanges would have occurred between the two groups, compared to the theoretical expectation of 16 trades (for details, see Knetsch and Sinden [1987]). This V/V^* ratio of .56 is quite similar to the ratios observed in experiments 1–4. In the study by Brookshire and Coursey (1987), the ratio of mean prices was indeed reduced by experience, from a high of 77 for initial hypothetical survey responses to 6.1 in the first potentially binding auction conducted in a laboratory. However, the ratio remained at 5.6 in the final auction.

3. Testing for Misrepresentation

As previously stated, subjects faced identical incentives in the induced-value and consumption goods phases of experiments 1–4. Therefore, it seems safe to attribute the difference in observed trading to the endowment effect. However, some readers of early drafts of this paper have suggested that because of the way market prices were determined, subjects might have felt that they had an incentive to misstate their true values in order to influence the price, and perhaps this incentive was perceived to be greater in the consumption goods markets. To eliminate this possible interpretation of the previous results, experiment 5 was carried out in a manner similar to the first four experiments, except that subjects were told that the price would be selected at random. As is well known, this is an incentive-compatible procedure for eliciting values (see Becker, DeGroot, and Marschak 1964).

Each participant received the following instructions (with appropriate alternative wording in the buyers' forms):

> After you have finished, one of the prices listed below will be selected at random and any exchanges will take place at that price. If you have indicated you will sell at this price you will receive this amount of money and will give up the mug; if you have indicated that you will keep the mug at this price then no exchange will be made and you can take the mug home with you.
>
> . . . Your decision can have no effect on the price actually used because the price will be selected at random.

[5] The ratio of the mean selling and buying prices is 1.4 if all subjects are included. However, if one buyer and one seller with extreme valuations are excluded, the ratio is 1.9. These numbers were reported in an earlier version of Coursey et al. (1987).

The experiment was conducted in a series of six tutorial groups of a business statistics class at Simon Fraser University. The use of small groups helped assure complete understanding of the instructions, and the exercises were conducted over the course of a single day to minimize opportunities for communication between participants. Each group was divided equally: half of the subjects were designated as sellers by random selection, and the other half became buyers. A total of 59 people took part.

Two induced-value markets for hypothetical payoffs and a subsequent third real exchange market for money and mugs were conducted with identical trading rules used in all three. All participants maintained the same role as either buyers or sellers for the three markets. As in experiments 1 and 2, the prices that individuals chose to buy or to sell were selected from possible prices ranging from $0.00 to $9.50 listed by increments of $0.50.

The results of this experiment were nearly identical to the earlier ones in which the actual exchanges were based on the market-clearing price. Even though possibly less motivating hypothetical values were used in the two induced-value markets, nearly all participants pursued a profit-maximizing selection of prices to buy or sell the assets. Fourteen exchanges at a price of $4.75 were expected in the first induced-value market on the basis of the randomly distributed values written on the forms. Thirteen trades at this price were indicated by the prices actually selected by the participants. The results of the second hypothetical induced-value market were equally convincing, with 16 of the 17 expected exchanges made at the expected price of $5.75. The procedures and incentives were apparently well understood by the participants.

Mugs, comparable to those used in other experiments, were distributed to the potential sellers after the induced-value markets were completed. A mug was also shown to all the potential buyers. The following form with instructions, nearly identical to the ones used in the induced-value markets, was then distributed (with the alternative wording for buyers in brackets):

> You now [do not] have, and own a mug which you can keep and take home. You also have the option of selling it and receiving [buying one to take home by paying] money for it.
>
> For each of the possible prices listed below, please indicate whether you wish to: (1) Receive [pay] that amount of money and sell your [buy a] mug, or (2) Not sell your [buy a] mug at this price.
>
> After you have finished, one of the prices listed below will be selected at random and any exchanges will take place at that price. If you have indicated you will sell [buy] at this price you will receive this amount of money [a mug] and will give up the mug [pay this amount of money]; if you have indicated that you will keep the [not buy a] mug at this price then no exchange will be made and you can take the mug home with you [do not pay anything].
>
> Notice the following two things: (1) Your decision can have no effect on the price actually used because the price will be selected at random. (2) It is in your interest to indicate your true preferences at each of the possible prices listed below.

For each price indicate your decision by marking an X in the appropriate column.

	I will sell [buy]	I will keep [not buy] the mug
If the price is $0	________	________
If the price is $0.50	________	________
.		
.		
.		
If the price is $9.50	________	________

After the instructions were read and reviewed by the experimenter and questions were answered, participants completed the forms indicating either their lowest selling price or their highest buying price. A random price, from among the list from $0.00 to $9.50, was then drawn, and exchanges based on this price were completed.

The results again showed a large and significant endowment effect. Given the 29 potential buyers, 30 potential sellers, and the random distribution of the mugs, 14.5 exchanges would be expected if entitlements did not influence valuations. Instead, only 6 were indicated on the basis of the values actually selected by the potential buyers and sellers ($V/V^* = .41$). The median selling price of $5.75 was over twice the median buying price of $2.25, and the means were $5.78 and $2.21, respectively.

4. Reluctance to Buy versus Reluctance to Sell

Exchanges of money and a good (or between two goods) offer the possibilities of four comparisons: a choice of gaining either the good or money, a choice of losing one or the other, buying (giving up money for the good), and selling (giving up the good for money) (Tversky and Kahneman, in press). The endowment effect results from a difference between the relative preferences for the good and money. The comparison of buying and selling to simple choices between gains permits an analysis of the discrepancy between WTA and WTP into two components: reluctance to sell (exchanging the good for money) and reluctance to buy (exchanging money for the good).

Experiments 6 and 7 were carried out to assess the weight of reluctance to buy and reluctance to sell in undertrading of a good similar to the goods used in the earlier experiments. The subjects in experiment 6 were 77 Simon Fraser students, randomly assigned to three groups. Members of one group, designated sellers, were given a coffee mug and were asked to indicate whether or not they would sell the mug at a series of prices ranging from $0.00 to $9.25. A group of buyers indicated whether they were willing to buy a mug at each of these prices. Finally, choosers were asked to choose, for each of the possible prices, between a mug and cash.

The results again reveal substantial undertrading: While 12.5 trades were expected between buyers and sellers, only three trades took place ($V/V^* = .24$). The median valuations were $7.12 for sellers, $3.12 for choosers, and $2.87 for buyers. The close similarity of results for buyers and choosers indicates that there was relatively little reluctance to pay for the mug.

Experiment 7 was carried out with 117 students at the University of British Columbia. It used an identical design except that price tags were left on the mugs. The results were consistent with those in experiment 6. Nineteen trades were expected on the basis of valuation equivalence, but only one was concluded on the basis of actual valuations ($V/V^* = .05$). The median valuations were $7.00 for sellers, $3.50 for choosers, and $2.00 for buyers.

It is worth noting that these results eliminate any form of income effect as an explanation of the discrepant valuations since the positions of sellers and choosers were strictly identical. The allocation of a particular mug to each seller evidently induced a sense of endowment that the choosers did not share: the median value of the mug to the sellers was more than double the value indicated by the choosers even though their choices were objectively the same. The results imply that the observed undertrading of consumption goods may be largely due to a reluctance to part with entitlements.

5. Bilateral Bargaining and the Coase Theorem

According to the Coase Theorem, the allocation of resources to individuals who can bargain and transact at no cost should be independent of initial property rights. However, if the marginal rate of substitution between one good and another is affected by endowment, then the individual who is assigned the property right to a good will be more likely to retain it. A bilateral bargaining experiment (experiment 8) was carried out to test this implication of the endowment effect.

The subjects were 35 pairs of students in 7 small tutorials at Simon Fraser University. The students were enrolled in either a beginning economics course or an English class. Each student was randomly paired with another student in the same tutorial group, with care taken to assure that students entering the tutorial together were not assigned as a pair. A game of Nim, a simple game easily explained, was played by each pair of participants. The winners of the game were each given a 400-gram Swiss chocolate bar and told it was theirs to keep.

An induced-value bargaining session was then conducted. The member of each pair who did not win the Nim game, and therefore did not receive the chocolate bar, was given a ticket and an instruction sheet that indicated that the ticket was worth $3.00 because it could be redeemed for that sum. The ticket owners were also told that they could sell the ticket to their partner if mutually agreeable terms could be reached. The partners (the chocolate bar owners) received instructions indicating that they could receive $5.00 for the ticket if they could successfully buy it from the owner. Thus there was a $2.00 surplus available to any pair completing a trade.

The pairs were then given an unlimited amount of time to bargain. Subjects were told that both credit and change were available from the experimenter. Results of the bargaining sessions were recorded on their instruction sheets.

Of the 35 pairs of participants, 29 agreed to an exchange ($V/V^* = .83$). The average price paid for the 29 tickets was $4.09, with 12 of the exchange prices being exactly $4.00. Payments of the redemption values of the tickets were made as soon as the exchanges were completed. These payments were made in single dollar bills to facilitate trading in the subsequent bargaining session. After the ticket exchanges were completed, owners of the chocolate bars were told that they could sell them to their partners if a mutually agreeable price could be determined. The procedures used for the tickets were once again applied to these bargaining sessions.

An important effect of the preliminary induced-value ticket bargains was to provide the ticket owners with some cash. The average gain to the ticket owners (including the six who did not sell their tickets) was $3.90. The average gain to their partners (the chocolate bar owners) was only $0.76. Thus the potential chocolate bar buyers were endowed with an average of $3.14 more than the owners, creating a small income effect toward the buyers. Also, to the extent that a windfall gain such as this is spent more casually by subjects than other money (for evidence on such a "house money effect," see Thaler and Johnson [1990]), trading of chocolate bars should be facilitated.

Results of the chocolate bar bargains once again suggest reluctance to trade. Rather than the 17.5 trades expected from the random allocations, only seven were observed ($V/V^* = .4$). The average price paid in those exchanges that did occur was $2.69 (the actual prices were $6.00, $3.10, $3.00, $2.75, $2.00, $1.00, and $1.00). If the six pairs of subjects who did not successfully complete bargains in the first stage are omitted from the sample on the grounds that they did not understand the task or procedures, then six trades are observed where 14.5 would be expected ($V/V^* = .414$). Similarly, if two more pairs are dropped because the prices at which they exchanged tickets were outside the range $3.00–$5.00, then the number of trades falls to four, and V/V^* falls to .296. (No significant differences between the students in the English and economics classes were observed.)[6]

To be sure that the chocolate bars were valued by the subjects and that these valuations would vary enough to yield mutually beneficial trades, the same chocolate bars were distributed to half the members of another class at Simon Fraser. Those who received chocolate bars were asked the minimum price they would accept to sell their bar, while those without the bars were asked the maximum price they would pay to acquire a bar. The valuations of the bars varied from $0.50 to $8.00. The average value ascribed by sellers was $3.98, while the buyers' average valuation was $1.25. (The median values were $3.50 and $1.25.)

[6] We conducted two similar bargaining experiments that yielded comparable results. Twenty-six pairs of subjects negotiated the sale of mugs and then envelopes containing an uncertain amount of money. Buyers had not been given any cash endowment. These sessions yielded 6 and 5 trades, respectively, where 13 would be expected. Also, some induced-value bilateral negotiation sessions were conducted in which only $0.50 of surplus was available (the seller's valuation was $1.50 and the buyer's was $2.00). Nevertheless, 21 of a possible 26 trades were completed.

6. The Endowment Effect in Choices between Goods

The previous experiments documented undertrading in exchanges of money and consumption goods. A separate experiment (Knetsch 1989) establishes the same effect in exchanges between two goods. Participants in three classes were offered a choice between the same two goods. All students in one class were given a coffee mug at the beginning of the session as compensation for completing a short questionnaire. At the completion of the task, the experimenters showed the students a bar of Swiss chocolate that they could immediately receive in exchange for the mug. The students in another class were offered an opportunity to make the opposite exchange after first being given the chocolate bar. The students in a third class were simply offered a choice, at the beginning of the session, between a chocolate bar and a mug. The proportion of students selecting the mug was 89 percent in the class originally endowed with mugs ($N = 76$), 56 percent in the class offered a choice ($N = 55$), and only 10 percent in the class originally endowed with chocolate bars ($N = 87$). For most participants a mug was more valuable than the chocolate when the mug had to be given up but less valuable when the chocolate had to be given up. This experiment confirms that undertrading can occur even when income effects are ruled out. It also demonstrates an endowment effect for a good that was distributed to everyone in the class and therefore did not have the appeal of a prize or trophy.

7. Discussion

The evidence presented in this chapter supports what may be called an instant endowment effect: the value that an individual assigns to such objects as mugs, pens, binoculars, and chocolate bars appears to increase substantially as soon as that individual is given the object.[7] The apparently instantaneous nature of the reference point shift and consequent value change induced by giving a person possession of a good goes beyond previous discussions of the endowment effect, which focused on goods that have been in the individual's possession for some time. While long-term endowment effects could be explained by sentimental attachment or by an improved technology of consumption in the Stigler-Becker (1977) sense, the differences in preference or taste demonstrated by more than 700 participants in the experiments reported in this paper cannot be explained in this fashion.

The endowment effect is one explanation for the systematic differences between buying and selling prices that have been observed so often in past work.

[7] The impression gained from informal pilot experiments is that the act of giving the participant physical possession of the good results in a more consistent endowment effect. Assigning subjects a chance to receive a good, or a property right to a good to be received at a later time, seemed to produce weaker effects.

One of the objectives of this study was to examine an alternative explanation for this buying-selling discrepancy, namely that it reflects a general bargaining strategy (Knez and Smith 1987) that would be eliminated by experience in the market (Brookshire and Coursey 1987; Coursey et al. 1987). Our results do not support this alternative view. The trading institution used in experiments 1–7 encouraged participants to be price takers (especially in experiment 5), and the rules provided no incentive to conceal true preferences. Furthermore, the results of the induced-value markets indicate that the subjects understood the demand-revealing nature of the questions they were asked and acted accordingly. Substantial undertrading was nevertheless observed in markets for consumption goods. As for learning and market discipline, there was no indication that buying and selling prices converged over repeated market trials, though full feedback was provided at the end of each trial. The undertrading observed in these experiments appears to reflect a true difference in preferences between the potential buyers and sellers. The robustness of this result reduces the risk that the outcome is produced by an experimental artifact. In short, the present findings indicate that the endowment effect can persist in genuine market settings.

The contrast between the induced-value markets and the consumption goods markets lends support to Heiner's (1985) conjecture that the results of induced-value experiments may not generalize to all market settings. The defining characteristic of the induced-value markets is that the values of the tokens are unequivocally defined by the amount the experimenter will pay for them. Loss-aversion is irrelevant with such objects because transactions are evaluated simply on the basis of net gain or loss. (If someone is offered \$6.00 for a \$5.00 bill, there is no sense of loss associated with the trade.) Some markets may share this feature of induced-value markets, especially when the conditions of pure arbitrage are approached. However, the computation of net gain and loss is not possible in other situations, for example, in markets in which risky prospects are traded for cash or in markets in which people sell goods that they also value for their use. In these conditions, the cancellation of the loss of the object against the dollars received is not possible because the good and money are not strictly commensurate. The valuation ambiguity produced by this lack of commensurability is necessary, although not sufficient, for both loss aversion and a buying-selling discrepancy.

The results of the experimental demonstrations of the endowment effect have direct implications for economic theory and economic predictions. Contrary to the assumptions of standard economic theory that preferences are independent of entitlements,[8] the evidence presented here indicates that people's preferences depend on their reference positions. Consequently, preference orderings are not defined independently of endowments: good A may be preferred to B when A is part of an original endowment, but the reverse may be true when initial reference positions are changed. Indifference curves will have a kink at the endowment

[8] Although ownership can affect taste in the manner suggested by Stigler and Becker (1977), in the absence of income effects, it is traditional to assume that the indifference curves in an Edgeworth box diagram do not depend on the location of the endowment point.

or reference point (see Tversky and Kahneman, in press), and an indifference curve tracing acceptable trades in one direction may even cross another indifference curve that plots the acceptable exchanges in the opposite direction (Knetsch 1989).

The existence of endowment effects reduces the gains from trade. In comparison with a world in which preferences are independent of endowment, the existence of loss-aversion produces an inertia in the economy because potential traders are more reluctant to trade than is conventionally assumed. This is not to say that Pareto-optimal trades will not take place. Rather, there are simply fewer mutually advantageous exchanges possible, and so the volume of trade is lower than it otherwise would be.

To assess the practical significance of the endowment effect, it is important to consider first some necessary conditions for the effect to be observed. Experiments 6 and 7 suggest that the endowment effect is primarily a problem for sellers; we observed little reluctance to buy but much reluctance to sell. Furthermore, not all sellers are afflicted by an endowment effect. The effect did not appear in the markets for money tokens, and there is no reason in general to expect reluctance to resell goods that are held especially for that purpose. An owner will not be reluctant to sell an item at a given price if a perfect substitute is readily available at a lower price. This reasoning suggests that endowment effects will almost certainly occur when owners are faced with an opportunity to sell an item purchased for use that is not easily replaceable. Examples might include tickets to a sold-out event, hunting licenses in limited supply (Bishop and Heberlein 1979), works of art, or a pleasant view.

While the conditions necessary for an endowment effect to be observed may appear to limit its applicability in economic settings, in fact these conditions are very often satisfied, and especially so in the bargaining contexts to which the Coase Theorem is applied. For example, tickets to Wimbledon are allocated by means of a lottery. A standard Coasean analysis would imply that in the presence of an efficient ticket brokerage market, winners of the lottery would be no more likely to attend the matches than other tennis fans who had won a similar cash prize in an unrelated lottery. In contrast, the experimental results presented in this chapter predict that many winners of Wimbledon tickets will attend the event, turning down opportunities to sell their tickets that exceed their reservation price for buying them.

Endowment effects can also be observed for firms and other organizations. Endowment effects are predicted for property rights acquired by historic accident or fortuitous circumstances, such as government licenses, landing rights, or transferable pollution permits. Owing to endowment effects, firms will be reluctant to divest themselves of divisions, plants, and product lines even though they would never consider buying the same assets; indeed, stock prices often rise when firms do give them up. Again, the prediction is not an absence of trade, just a reduction in the volume of trade.

Isolating the influence of endowment effects from those of transaction costs as causes of low trading volumes is, of course, difficult in actual market settings.

Demonstrations of endowment effects are most persuasive where transaction costs are very small. By design, this was the case in the experimental markets, where the efficiency of the induced-value markets demonstrated the minimal effect of transaction costs, or other impediments, on exchange decisions, leaving the great reluctance to trade mugs and other goods to be attributable to endowment effects.

Endowment effects are not limited to cases involving physical goods or to legal entitlements. The reference position of individuals and firms often includes terms of previous transactions or expectations of continuation of present, often informal, arrangements. There is clear evidence of dramatically asymmetric reactions to improvements and deteriorations of these terms and a willingness to make sacrifices to avoid unfair treatment (Kahneman, Knetsch, and Thaler 1986). The reluctance to sell at a loss, owing to a perceived entitlement to a formerly prevailing price, can explain two observations of apparent undertrading. The first pertains to housing markets. It is often observed that when housing prices fall, volume also falls. When house prices are falling, houses remain on the market longer than when prices are rising. Similarly, the volume for stocks that have declined in price is lower than the volume for stocks that have increased in value (Shefrin and Statman 1985; Ferris, Haugen, and Makhija 1988), although tax considerations would lead to the opposite prediction.

Another manifestation of loss aversion in the context of multiattribute negotiations is what might be termed "concession-aversion": a reluctance to accept a loss on any dimension of an agreement. A straightforward and common instance of this is the downward stickiness of wages. A somewhat more subtle implication of concession aversion is that it can produce inefficient contract terms owing to historic precedents. Old firms may have more inefficient arrangements than new ones because new companies can negotiate without the reference positions created by prior agreements. Some airlines, for example, are required to carry three pilots on some planes while others—newer ones—operate with two.

Loss-aversion implies a marked asymmetry in the treatment of losses and forgone gains, which plays an essential role in judgments of fairness (Kahneman et al. 1986). Accordingly, disputes in which concessions are viewed as losses are often much less tractable than disputes in which concessions involve forgone gains. Court decisions recognize the asymmetry of losses and forgone gains by favoring possessors of goods over other claimants, by limiting recovery of lost profits relative to compensation for actual expenditures, and by failing to enforce gratuitous promises that are coded as forgone gains to the injured party (Cohen and Knetsch 1989).

To conclude, the evidence reported here offers no support for the contention that observations of loss aversion and the consequential evaluation disparities are artifacts; nor should they be interpreted as mistakes likely to be eliminated by experience, training, or "market discipline." Instead, the findings support an alternative view of endowment effects and loss-aversion as fundamental characteristics of preferences.

References

Banford, Nancy D., Jack L. Knetsch, and Gary A. Mauser. 1979. "Feasibility Judgements and Alternative Measures of Benefits and Costs." *Journal of Business Administration*, 11(1, 2): 25–35.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marschak. 1964. "Measuring Utility by a Single-Response Sequential Method." *Behavioral Science*, 9(July): 226–32.

Bishop, Richard C., and Thomas A. Heberlein. 1979. "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics*, 61(December): 926–30.

Boyce, Rebecca R., Thomas C. Brown, Gary D. McClelland, George L. Peterson, and William D. Schulze. 1990. "An Experimental Examination of Intrinsic Environmental Values." Working paper. Univ. Colorado, Boulder.

Brookshire, David S., and Don L. Coursey. 1987. "Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures." *American Economic Review*, 77 (September): 554–66.

Brookshire, David S., Alan Randall, and John R. Stoll. 1980. "Valuing Increments and Decrements in Natural Resource Service Flows." *American Journal of Agricultural Economics*, 62 (August): 478–88.

Cohen, David, and Jack L. Knetsch. 1989. "Judicial Choice and Disparities between Measures of Economic Values." Working paper. Simon Fraser Univ., Burnaby, B.C.

Coursey, Don L., John L. Hovis, and William D. Schulze. 1987. "The Disparity between Willingness to Accept and Willingness to Pay Measures of Value." *Quarterly Journal of Economics*, 102 (August): 679–90.

Ferris, Stephen P., Robert A. Haugen, and Anil K. Makhija. 1988. "Predicting Contemporary Volume with Historic Volume at Differential Price Levels: Evidence Supporting the Disposition Effect." *Journal of Finance*, 43 (July): 677–97.

Freeman, A. Myrick. 1979. *The Benefits of Environmental Improvement*. Washington: Resources for the Future.

Hammack, Judd, and Gardner Mallard, Brown, Jr. 1974. *Waterfowl and Wetlands: Toward Bio-economic Analysis*. Baltimore: Johns Hopkins Press (for Resources for the Future).

Heberlein, Thomas A., and Richard C. Bishop. 1985. "Assessing the Validity of Contingent Valuation: Three Field Experiments." Paper presented at the International Conference on Man's Role in Changing the Global Environment, Italy.

Heiner, Ronald A. 1985. "Experimental Economics: Comment." *American Economic Review*, 75 (March): 260–63.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler. 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review*, 76 (September): 728–41.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47 (March): 263–91.

Knetsch, Jack L. 1989. "The Endowment Effect and Evidence of Nonreversible Indifference Curves." *American Economic Review*, 79 (December): 1277–84.

Knetsch, Jack 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 (August): 507–21.

———. 1987. “The Persistence of Evaluation Disparities.” *Quarterly Journal of Economics*, 102 (August): 691–95.

Knez, Marc, and Smith, Vernon L. 1987. “Hypothetical Valuations and Preference Reversals in the Context of Asset Trading.” In *Laboratory Experiments in Economics: Six Points of View*, edited by Alvin E. Roth. Cambridge: Cambridge University Press.

Knez, Peter, Vernon L. Smith, and Arlington W. Williams. 1985. “Individual Rationality, Market Rationality, and Value Estimation.” *American Economic Review, Papers and Proceedings*, 75 (May): 397–402.

Plott, Charles R. 1982. “Industrial Organization Theory and Experimental Economics.” *Journal of Economic Literature*, 20 (December): 1485–1527.

Rowe, Robert D., Ralph C. d’Arge, and David S. Brookshire. 1980. “An Experiment on the Economic Value of Visibility.” *Journal of Environmental Economy and Management*, 7 (March): 1–19.

Shefrin, Hersh, and Meil Statman. 1985. “The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence.” *Journal of Finance*, 40 (July): 777–90.

Sinclair, William F. 1978. *The Economic and Social Impact of Kemano II Hydroelectric Project on British Columbia’s Fisheries Resources*. Vancouver: Dept. Fisheries and Oceans.

Smith, Vernon L. 1976. “Experimental Economics: Induced Value Theory.” *American Economic Review*, *Papers and Proceedings*, 66 (May): 274–79.

———. 1982. “Macroeconomic Systems as an Experimental Science.” *American Economic Review*, 72 (December): 923–55.

Stigler, George J., and Becker, Gary S. 1977. “De Gustibus Non Est Disputandum.” *American Economic Review*, 67 (March): 76–90.

Thaler, Richard. 1980. “Toward a Positive Theory of Consumer Choice.” *Journal of Economic Behavior and Organization*, 1 (March): 39–60.

Thaler, Richard, and Eric J. Johnson. 1990. “Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice.” *Management Science*, 36 (June).

Tversky, Amos, and Daniel Kahneman. 1991. “Loss Aversion in Riskless Choice: A Reference-Dependent Model.” *Quarterly Journal of Economics*, 106: 1039–61.

Willig, Robert D. 1976. “Consumer’s Surplus without Apology.” *American Economic Review*, 66 (September): 589–97.

C H A P T E R 3

Mental Accounting Matters

RICHARD H. THALER

• A former colleague of mine, a professor of finance, prides himself on being a thoroughly rational man. Long ago he adopted a clever strategy to deal with life's misfortunes. At the beginning of each year he establishes a target donation to the local United Way charity. Then, if anything untoward happens to him during the year, for example an undeserved speeding ticket, he simply deducts this loss from the United Way account. He thinks of it as an insurance policy against small annoyances.[1]

• A few years ago I gave a talk to a group of executives in Switzerland. After the conference my wife and I spent a week visiting the area. At that time the Swiss franc was at an all-time high relative to the US dollar, so the usual high prices in Switzerland were astronomical. My wife and I comforted ourselves that I had received a fee for the talk that would easily cover the outrageous prices for hotels and meals. Had I received the same fee a week earlier for a talk in New York though, the vacation would have been much less enjoyable.

• A friend of mine was once shopping for a quilted bedspread. She went to a department store and was pleased to find a model she liked on sale. The spreads came in three sizes: double, queen and king. The usual prices for these quilts were $200, $250 and $300 respectively, but during the sale they were all priced at only $150. My friend bought the king-size quilt and was quite pleased with her purchase, though the quilt did hang a bit over the sides of her double bed.

Introduction

The preceding anecdotes all illustrate the cognitive processes called mental accounting. What is mental accounting? Perhaps the easiest way to define it is to compare it with financial and managerial accounting as practised by organizations.

I have been thinking about mental accounting for more than twenty years, so it is not possible to thank everyone who has helped me write this chapter. Some who have helped recently include John Gourville, Chip Heath, Daniel Kahneman, France Leclerc, George Loewenstein, Cade Massey, Drazen Prelec, Dilip Soman, and Roman Weil. This chapter began as an invited lecture to the SPUDM conference in Aix-en-Provence held in 1993. It was finally completed during my stay at The Center for Advanced Study in the Behavioral Sciences. Their help in reaching closure is gratefully acknowledged.

[1] This strategy need not reduce his annual contribution to the United Way. If he makes his intended contribution too low he risks having 'uninsured' losses. So far he has not been 'charitable' enough to have this fund cover large losses, such as when a hurricane blew the roof off his beach house.

According to my dictionary accounting is "the system of recording and summarizing business and financial transactions in books, and analyzing, verifying, and reporting the results." Of course, individuals and households also need to record, summarize, analyze, and report the results of transactions and other financial events. They do so for reasons similar to those that motivate organizations to use managerial accounting: to keep trace of where their money is going, and to keep spending under control. Mental accounting is a description of the ways they do these things.

How *do* people perform mental accounting operations? Regular accounting consists of numerous rules and conventions that have been codified over the years. You can look them up in a textbook. Unfortunately, there is no equivalent source for the conventions of mental accounting; we can learn about them only by observing behavior and inferring the rules.

Three components of mental accounting receive the most attention here. The first captures how outcomes are perceived and experienced, and how decisions are made and subsequently evaluated. The accounting system provides the inputs to do both ex ante and ex post cost—benefit analyses. This component is illustrated by the anecdote above involving the purchase of the quilt. The consumer's choice can be understood by incorporating the value of the "deal" (termed transaction utility) into the purchase decision calculus.

A second component of mental accounting involves the assignment of activities to specific accounts. Both the sources and uses of funds are labeled in real as well as in mental accounting systems. Expenditures are grouped into categories (housing, food, etc.) and spending is sometimes constrained by implicit or explicit budgets. Funds to spend are also labeled, both as flows (regular income versus windfalls) and as stocks (cash on hand, home equity, pension wealth, etc.). The first two anecdotes illustrate aspects of this categorization process. The vacation in Switzerland was made less painful because of the possibility of setting up a Swiss lecture mental account, from which the expenditures could be deducted. Similarly, the notional United Way mental account is a flexible way of making losses less painful.

The third component of mental accounting concerns the frequency with which accounts are evaluated and what Read, Loewenstein, and Rabin (1998) have labeled "choice bracketing." Accounts can be balanced daily, weekly, yearly, and so on, and can be defined narrowly or broadly. A well-known song implores poker players to "never count your money while you're sitting at the table." An analysis of dynamic mental accounting shows why this is excellent advice, in poker as well as in other situations involving decision making under uncertainty (such as investing).

The primary reason for studying mental accounting is to enhance our understanding of the psychology of choice. In general, understanding mental accounting processes helps us understand choice because mental accounting rules are not neutral.[2] That is, accounting decisions such as to which category to assign a

[2] An accounting system is a way of aggregating and summarizing large amounts of data to facilitate good decision making. In an ideal world the accounting system would accomplish this task in such a way that the decision maker would make the same choice when presented with only the accounting

purchase, whether to combine an outcome with others in that category, and how often to balance the 'books' can affect the perceived attractiveness of choices. They do so because mental accounting violates the economic notion of fungibility. Money in one mental account is not a perfect substitute for money in another account. Because of violations of fungibility, mental accounting matters.

The goal of this paper is to illustrate how mental accounting matters. To this end I draw upon research conducted over the past two decades. This describes where I think the field is now, having been informed by the research of many others, especially over the past few years.

The Framing of Gains and Losses

The Value Function

We wish to understand the decision-making process of an individual or a household interacting in an economic environment. How does a person make economic decisions, such as what to buy, how much to save, and whether to buy or lease an item? And how are the outcomes of these financial transactions evaluated and experienced?

Following my earlier treatment of these questions (Thaler 1980, 1985) I assume that people perceive outcomes in terms of the value function of Kahneman and Tversky's (1979) prospect theory. The value function can be thought of as a representation of some central components of the human perceived pleasure machine.[3] It has three important features, each of which captures an essential element of mental accounting:

1. *The value function is defined over gains and losses relative to some reference point*. The focus on changes, rather than wealth levels as in expected utility theory, reflects the piecemeal nature of mental accounting. Transactions are often evaluated one at a time, rather than in conjunction with everything else.

2. *Both the gain and loss functions display diminishing sensitivity*. That is, the gain function is concave and the loss function is convex. This feature reflects the basic psychophysical principle (the Weber-Fechner law) that the difference between $10 and $20 seems bigger than the difference between $1000 and $1010, irrespective of the sign.

3. *Loss-aversion*. Losing $100 hurts more than gaining $100 yields pleasure: $v(x) < -v(-x)$. The influence of loss aversion on mental accounting is enormous, as will become evident very quickly.

---

data as she would if she had access to all the relevant data. This is what I mean by "neutral." In a sense, such an accounting system would provide decision makers with "sufficient statistics." Of course, achieving this goal is generally impossible because something must be sacrificed in order to reduce the information the decision maker has to look at. Thus neither organizational nor mental accounting will achieve neutrality.

[3] Prospect theory predates Kahneman's (1994) important distinction between decision utility and experienced utility. In his terms, the prospect theory value function measures decision utility.

Decision Frames

The role of the value function in mental accounting is to describe how events are perceived and coded in making decisions. To introduce this topic, it is useful to define some terms. Tversky and Kahneman (1981, p. 456) define a mental account[4] quite narrowly as "an outcome frame which specifies (i) the set of elementary outcomes that are evaluated jointly and the manner in which they are combined and (ii) a reference outcome that is considered neutral or normal." (Typically, the reference point is the status quo.) According to this definition, a mental account is a frame for evaluation. I wish to use the term 'mental accounting' to describe the entire process of coding, categorizing, and evaluating events, so this narrow definition of a mental account is a bit confining. Accordingly, I will refer to simply outcome frames as "entries."

In a later paper, Kahneman and Tversky (1984, p. 347), propose three ways that outcomes might be framed: in terms of a minimal account, a topical account, or a comprehensive account. Comparing two options using the minimal account entails examining only the differences between the two options, disregarding all their common features. A topical account relates the consequences of possible choices to a reference level that is determined by the context within which the decision arises. A comprehensive account incorporates all other factors including current wealth, future earnings, possible outcomes of other probabilistic holdings, and so on. (Economic theory generally assumes that people make decisions using the comprehensive account.) The following example[5] illustrates that mental accounting is topical:

> Imagine that you are about to purchase a jacket for ($125)[$15] and a calculator for ($15)[$125]. The calculator salesman informs you that the calculator you wish to buy is on sale for ($10)[$120] at the other branch of the store, located 20 minutes drive away. Would you make the trip to the other store?[5] (Tversky and Kahneman 1981, p. 459)

When two versions of this problem are given (one with the figures in parentheses, the other with the figures in brackets), most people say that they will travel to save the $5 when the item costs $15 but not when it costs $125. If people were using a minimal account frame they would be just asking themselves whether they are willing to drive 20 minutes to save $5, and would give the same answer in either version.

Interestingly, a similar analysis applies in the comprehensive account frame. Let existing wealth be W, and W^* be existing wealth plus the jacket and calculator minus $140. Then the choice comes down to the utility of W^* plus $5 versus the utility of W^* plus 20 minutes. This example illustrates an important general

[4] Actually, they use the term "psychological account" in their 1981 paper, following the terminology I used in my 1980 paper. Later (Kahneman and Tversky 1984) they suggest the better term "mental account."

[5] This problem was based on similar examples discussed by Savage (1954) and Thaler (1980).

point—the way a decision is framed will not alter choices if the decision maker is using a comprehensive, wealth-based analysis. Framing does alter choices in the real world because people make decisions piecemeal, influenced by the context of the choice.

Hedonic Framing

The jacket and calculator problem does demonstrate that mental accounting is piecemeal and topical, but there is more to learn from this example. Why are we more willing to drive across town to save money on a small purchase than a large one? Clearly there is some psychophysics at work here. Five dollars seems like a significant saving on a $15 purchase, but not so on a $125 purchase. But this disparity implies that the utility of the saving must be associated with the differences in values rather than the value of the difference. That is, the utility of saving $5 on the purchase of the expensive item must be ($v(-\$125) - v(-\$120)$) (or perhaps the ratio of these values) rather than $v(\$5)$, otherwise there would be no difference between the two versions of the problem.

What else do we know about mental accounting arithmetic? Specifically, how are two or more financial outcomes (within a single account) combined? This is an important question because we would like to be able to construct a model of how consumers evaluate events such as purchases that typically involve combinations of outcomes, good or bad.

One possible place to start in building a model of how people code combinations of events is to assume they do so to make themselves as happy as possible. To characterize this process we need to know how someone with a prospect theory value function could wish to have the receipt of multiple outcomes framed. That it, for two outcomes x and y, when will $v(x + y)$ be greater than $v(x) + v(y)$? I have previously considered this question (Thaler 1985). Given the shape of the value function, it is easy to derive the following principles of hedonic framing, that is, the way of evaluating joint outcomes to maximize utility:

1. Segregate gains (because the gain function is concave).
2. Integrate losses (because the loss function is convex).
3. Integrate smaller losses with larger gains (to offset loss aversion).
4. Segregate small gains (silver linings) from larger losses (because the gain function is steepest at the origin, the utility of a small gain can exceed the utility of slightly reducing a large loss).

As I showed, most people share the intuition that leads to these principles. That is, if you ask subjects "Who is happier, someone who wins two lotteries that pay $50 and $25 respectively, or someone who wins a single lottery paying $75?" Sixty-four percent say the two-time winner is happier. A similar majority shared the intuition of the other three principles.

These principles are quite useful in thinking about marketing issues. In other words, if one wants to describe the advantages and disadvantages of a particular product in a way that will maximize the perceived attractiveness of the product to consumers, the principles of hedonic framing are a helpful guide. For example, framing a sale as a "rebate" rather than a temporary price reduction might facilitate the segregation of the gain in line with principle 4.

The Failure of the Hedonic Editing Hypothesis

It would be convenient if these same principles could also serve as a good descriptive model of mental accounting. Can people be said to edit or parse the multiple outcomes they consider or experience in a way that could be considered optimal, that is, hedonic editing.[6] More formally, if the symbol "&" is used to denote the cognitive combination of two outcomes, then hedonic editing is the application of the following rule:

$$v(x \,\&\, y) = \mathrm{Max}[v(x+y), v(x)+v(y)]$$

The hypothesis that people engage in hedonic editing has obvious theoretical appeal[7] but some thought reveals that it cannot be descriptively correct. Consider the jacket and calculator problem again. If the $5 saving were coded in a utility-maximizing way it would be segregated in either case, inconsistent with the data. Furthermore, there must be some limits to our abilities to engage in self-deception. Why stop at segregating the $5 gain? Why not code it as five gains of $1? Nevertheless, hedonic editing represents a nice starting point for the investigation of how people do code multiple events.

Eric Johnson and I have investigated the limits of the hedonic editing hypothesis (Thaler and Johnson, 1990). Our ultimate goal was to explore the influence of prior outcomes on risky choices (see below), but we began with the more basic question of how people choose to code multiple events such as a gain of $30 followed by a loss of $9. One approach we used was to ask people their preferences about temporal spacing. For two specified financial outcomes, we asked subjects who would be happier, someone who had these two events occur on the same day, or a week or two apart? The reasoning for this line of inquiry was that temporal separation would facilitate cognitive segregation. So if a subject wanted to segregate the outcomes x and y, he would prefer to have them occur on different days, whereas if he wanted to integrate them, he would prefer to have them occur together. The hedonic editing hypothesis would be supported if subjects preferred temporal separation for cases where the hypothesis called for segregation, and temporal proximity when integration was preferred. For gains, the hedonic editing hypothesis was supported. A large majority of subjects thought temporal separation of gains produced more happiness. But, in contrast to the hedonic editing hypothesis, subjects thought separating losses was also a good idea. Why?

[6] Johnson and I used the term 'editing' for this process, though on reflection 'parsing' might have been better. I will stick with the original term to avoid confusion with the prior literature. Note that editing refers to active cognitions undertaken by the decision maker. In contrast, I will use 'framing' to refer to the way a problem is posed externally. As we will see, people prefer to have outcomes framed hedonically, but fail to edit (or one could say, reframe) them accordingly.

[7] See Fishburn and Luce (1995) for an axiomatic treatment of hedonic editing.

The intuition for the hypothesis that people would want to combine losses comes from the fact that the loss function displays diminishing sensitivity. Adding one loss to another should diminish its marginal impact. By wishing to spread out losses, subjects seem to be suggesting that they think that a prior loss makes them *more* sensitive towards subsequent losses, rather than the other way around. In other words, subjects are telling us that they are unable to simply add one loss to another (inside the value function parentheses). Instead, they feel that losses must be felt one by one, and that bearing one loss makes one more sensitive to the next.[8]

To summarize, the evidence suggests that the rules of hedonic framing are good descriptions of the way people would like to have the world organized (many small gains including silver linings; losses avoided if possible but otherwise combined). People will also actively parse outcomes consistent with these rules, with the exception of multiple losses.

There are two important implications of these results for mental accounting. First, we would expect mental accounting to be as hedonically efficient as possible. For example, we should expect that opportunities to combine losses with larger gains will be exploited wherever feasible. Second, loss aversion is even more important than the prospect theory value function would suggest, as it is difficult to combine losses to diminish their impact. This result suggests that we should expect to see that some of the discretion inherent in any accounting system will be used to avoid having to experience losses.

Mental Accounting Decision-Making

Transaction utility

What happens when a consumer decides to buy something, trading money for some object? One possibility would be to code the acquisition of the product as a gain and the forgone money as a loss. But loss aversion makes this frame hedonically inefficient. Consider a thirsty consumer who would rather have a can of soda than one dollar and is standing in front of a vending machine that sells soda for 75 cents. Clearly the purchase makes her better off, but it might be rejected if the payment were cognitively multiplied by 2.25 (an estimate of the coefficient of loss-aversion). This thinking has led both Kahneman and Tversky (1984) and me (Thaler 1985) to reject the idea that costs are generally viewed as losses.

Instead, I proposed that consumers get two kinds of utility from a purchase: *acquisition utility* and *transaction utility*. Acquisition utility is a measure of the value of the good obtained relative to its price, similar to the economic concept of consumer surplus. Conceptually, acquisition utility is the value the consumer would place on receiving the good as a gift, minus the price paid. Transaction utility

[8] Linville and Fischer (1991) also investigate the predictive power of hedonic editing, with similar results.

measures the perceived value of the "deal." It is defined as the difference between the amount paid and the 'reference price' for the good, that is, the regular price that the consumer expects to pay for this product. The following example (from Thaler, 1985) illustrates the role of transaction utility.

> You are lying on the beach on a hot day. All you have to drink is ice water. For the last hour you have been thinking about how much you would enjoy a nice cold bottle of your favorite brand of beer. A companion gets up to go make a phone call and offers to bring back a beer from the only nearby place where beer is sold (a fancy resort hotel) [a small, run-down grocery store]. He says that the beer might be expensive and so asks how much you are willing to pay for the beer. He says that he will buy the beer if it costs as much or less than the price you state. But if it costs more than the price you state he will not buy it. You trust your friend, and there is no possibility of bargaining with the (bartender) [store owner]. What price do you tell him?

Two versions of the question were administered, one using the phrases in parentheses, the other the phrases in brackets. The median responses for the two versions were $2.65 (resort) and $1.50 [store] in 1984 dollars. People are willing to pay more for the beer from the resort because the reference price in that context is higher. Note that this effect cannot be accommodated in a standard economic model because the consumption experience is the same in either case; the place of purchase should be irrelevant.

The addition of transaction utility to the purchase calculus leads to two kinds of effects in the marketplace. First, some goods are purchased primarily because they are especially good deals. Most of us have some rarely worn items in our closets that are testimony to this phenomenon. Sellers make use of this penchant by emphasizing the savings relative to the regular retail price (which serves as the suggested reference price). In contrast, some purchases that would seemingly make the consumer better off may be avoided because of substantial negative transaction utility. The thirsty beer-drinker who would pay $4 for a beer from a resort but only $2 from a grocery store will miss out on some pleasant drinking when faced with a grocery store charging $2.50.

Opening and Closing Accounts

One of the discretionary components of an accounting system is the decision of when to leave accounts 'open' and when to 'close' them. Consider the example of someone who buys 100 shares of stock at $10 a share. This investment is initially worth $1000, but the value will go up or down with the price of the stock. If the price changes, the investor has a "paper" gain or loss until the stock is sold, at which point the paper gain or loss becomes a 'realized' gain or loss. The mental accounting of paper gains and losses is tricky (and depends on timing—see below), but one clear intuition is that a realized loss is more painful than a paper loss. When a stock is sold, the gain or loss has to be "declared" both to the tax authorities and to the investor (and spouse). Because closing an account at a loss is painful, a prediction of mental accounting is that people will be reluctant to sell securities that have declined in value. In particular, suppose an investor needs to raise some cash and must choose between two stocks to sell, one of which has increased in value and one of which has decreased. Mental accounting favors selling the winner (Shefrin and Statman 1987) whereas a rational analysis favors selling the loser.[9] Odean (1998) finds strong support for the mental accounting prediction. Using a data set that tracked the trades of investors using a large discount brokerage firm, Odean finds that investors were more likely to sell one of their stocks that had increased in value than one of their stocks that had decreased.[10]

Other evidence of a reluctance to close an account in the "red" comes from the world of real accounting. Most public corporations make official earnings announcements every quarter. Although earnings are audited, firms retain some discretion in how quickly to count various components of revenues and expenses, leaving them with some control over the actual number they report. Several recent papers (e.g., Burgstahler and Dichev 1997; Degeorge, Patel and Zeckhauser, forthcoming) show that firms use this discretionary power to avoid announcing earnings decreases and losses. Specifically, a plot of earnings per share (in cents per share) or change in earnings per share (this quarter versus same quarter last year) shows a sharp discontinuity at zero. Firms are much more likely to make a penny a share than to lose a penny a share, and are much more likely to exceed last year's earnings by a penny than to miss by a penny. So small losses are converted into small gains. In contrast, large gains seem to be trimmed down (to increase the chance of an increase again next year) whereas moderate losses are somewhat inflated (a procedure known in accounting circles as "taking the big bath"). Apparently, firms believe that shareholders (or potential shareholders) react to earnings announcements in a manner consistent with prospect theory.

Advance Purchases, Sunk Costs, and Payment Depreciation

Another situation in which a consumer has to decide when to open and close an account is when a purchase is made well in advance of consumption. Consider paying $100 for two tickets to a basketball game to be held in a month's time. Suppose that the tickets are being sold at the reference price so transaction utility is zero. In this case the consumer can be said to open an account at the point at which the tickets are purchased. At this time the account has a negative balance of $100. Once the date of the game comes and the game is attended, the account can be closed.

What happens if something (a blizzard) prevents the consumer from attending the game? In this case the consumer has to close the account at a loss of $100; in accounting terminology the loss has to be recognized. Notice that this event turns

[9] A rational investor will choose to sell the loser because capital gains are taxable and capital losses are deductible.

[10] Of course, such a strategy could be rational if the losers they kept subsequently increased in value more than the winners they sold, but this outcome was not observed. Indeed, these investors are not particularly savvy. The stocks they sell subsequently outperform the stocks they buy!

a cost into a loss, which is aversive. Still, why does the prior expenditure (now a sunk cost) makes someone more willing to go to the game in a blizzard (as in the example in Thaler 1980)?

To answer this question we need to consider how transactions are evaluated. For most routine purchases there is no ex post evaluation of the purchase when the account is closed. Such evaluations become more likely as the size of the transaction increases or as the purchase or situation becomes more unusual. Failing to attend an event that has been paid for makes the purchase highly salient and an evaluation necessary. By driving through the storm, the consumer can put the game back into the category of normal transactions that are not explicitly evaluated and thus avoid adding up the costs and benefits (barring an accident!). Furthermore, even if an ex post evaluation is made, the extra cost of going to the game may not be included in the evaluation. As Heath (1995) suggests, because the costs of driving to the game are not monetary, they may not be included in the analysis.[11] In Heath's terms they are incidental, that is, in a different mental account. He makes the telling comparison between this case and the Kahneman and Tversky (1984) theater-ticket example, in which subjects are *less* willing to buy a ticket to a play after having lost their ticket than after having lost an equivalent sum of money. In the theater-ticket example, buying a second ticket is aversive because it *is* included in the mental account for the theater outing, but the loss of the money is not.

Although sunk costs influence subsequent decisions, they do not linger indefinitely. A thought experiment illustrates this point nicely. Suppose you buy a pair of shoes. They feel perfectly comfortable in the store, but the first day you wear them they hurt. A few days later you try them again, but they hurt even more than the first time. What happens now? My predictions are as follows:

1. The more you paid for the shoes, the more times you will try to wear them. (This choice may be rational, especially if they have to be replaced with another expensive pair.)
2. Eventually you stop wearing the shoes, *but you do not throw them away*. The more you paid for the shoes, the longer they sit in the back of your closet before you throw them away. (This behavior cannot be rational unless expensive shoes take up less space.)
3. At some point, you throw the shoes away, regardless of what they cost, the payment having been fully "depreciated."

Evidence about the persistence of sunk costs effects is reported by Arkes and Blumer (1985). They ran an experiment in which people who were ready to buy season tickets to a campus theater group were randomly placed into three groups: one group paid full price, one group got a small (13%) discount, and one group

[11] Of course, although the driving costs may not be included in the basketball game account, they must be compared, at least prospectively, to something when one is deciding whether to go. In this formulation someone would choose to take the drive, not in order to enjoy the game, but to avoid feeling the pain associated with the unamortized ticket expense.

received a large (47%) discount. The experimenters then monitored how often the subjects attended plays during the season. In the first half of the season, those who paid full price attended significantly more plays than those who received discounts, but in the second half of the season there was no difference among the groups. People do ignore sunk costs, eventually.

The gradual reduction in the relevance of prior expenditures is dubbed "payment depreciation" by Gourville and Soman (1998), who have conducted a clever field experiment to illustrate the idea. They obtained usage data from the members of a health club that charges the dues to its members twice a year. Gourville and Soman find that attendance at the health club is highest in the month in which the dues are paid and then declines over the next five months, only to jump again when the next bill comes out.

Similar issues are involved in the mental accounting of wine collectors who often buy wine with the intention of storing it for ten years or more while it matures. When a bottle is later consumed, what happens? Eldar Shafir and I (1998) have investigated this pressing issue by surveying the subscribers to a wine newsletter aimed at serious wine consumers/collectors. We asked the following question:

> Suppose you bought a case of a good 1982 Bordeaux in the futures market for $20 a bottle. The wine now sells at auction for about $75 a bottle. You have decided to drink a bottle. Which of the following best captures your feeling of the cost to you of drinking this bottle?

We gave the respondents five answers to choose from: $0, $20, $20 plus interest, $75, and −$55 ("I drink a $75 bottle for which I paid only $20"). The percentages of respondents choosing each answer were 30, 18, 7, 20 and 25. Most of the respondents who selected the economically correct answer ($75) were in fact economists. (The newsletter, *Liquid Assets*, is published by economist Orley Ashenfelter and has many economist subscribers). More than half the respondents report that drinking the bottle either costs nothing or actually saves them money!

The results of this survey prompted us to run a follow-up survey the following year. The question this time was

> Suppose you buy a case of Bordeaux futures at $400 a case. The wine will retail at about $500 a case when it is shipped. You do not intend to start drinking this wine for a decade. At the time that you acquire this wine which statement more accurately captures your feelings?
>
> a. I feel like I just spent $400, much as I would feel if I spent $400 on a weekend getaway.
> b. I feel like I made a $400 investment that I will gradually consume after a period of years.
> c. I feel like I just saved $100, the difference between what the futures cost and what the wine will sell for when delivered.

Respondents rated each answer on a five-point scale. Most respondents selected answer (b) as their favorite, coding the initial purchase as an investment.

Notice that this choice means that the typical wine connoisseur thinks of his initial purchase as an investment and later thinks of the wine as free when he drinks it. We therefore titled our paper “Invest Now, Drink Later, Spend Never.” Note that this mental accounting transforms a very expensive hobby into one that is “free.” The same mental accounting applies to time-share vacation properties. The initial purchase of a week every year at some resort feels like an investment, and the subsequent visits feel free.

Payment Decoupling

In the wine example, the prepayment separates or “decouples” (Prelec and Loewenstein 1998; Gourvile and Soman 1998) the purchase from the consumption and in so doing seems to reduce the perceived cost of the activity. Prepayment can often serve this role, but the mental accounting advantages of decoupling are not all associated with prepayment. Consider the case of the pricing policies of the Club Med resorts (Thaler 1980). At these vacation spots consumers pay a fixed fee for a vacation that includes meals, lodging, and recreation. This plan has two advantages. First, the extra cost of including the meals and recreation in the price will look relatively small when combined with the other costs of the vacation. Second, under the alternative plan each of the small expenditures looks large by itself, and is likely to be accompanied by a substantial dose of negative transaction utility given the prices found at most resorts.

Another disadvantage of the piece-rate pricing policy is that it makes the link between the payment and the specific consumption act very salient, when the opposite is highly desirable. For example, a *prix fixe* dinner, especially an expensive multicourse meal, avoids the unsavory prospect of matching a very high price with the very small quantity of food offered in each course.[12] Along the same lines, many urban car owners would be financially better off selling their car and using a combination of taxis and car rentals. However, paying $10 to take a taxi to the supermarket or a movie is both salient and linked to the consumption act; it seems to raise the price of groceries and movies in a way that monthly car payments (or even better, a paid-off car) do not.

More generally, consumers don’t like the experience of “having the meter running.” This contributes to what has been called the “flat rate bias” in telecommunications. Most telephone customers elect a flat rate service even though paying by the call would cost them less.[13] Train (1991, p. 211) says that “consumers seem to value flat-rate service over measured service even when the bill that the consumer would receive under the two services, given the number of calls the consumer places, would be the same. . . . The existence of this bias is problematical.

[12] In contrast, the review of one expensive San Francisco restaurant in the Zagat guide includes the following gripe from a customer. ‘$13 for two scallops, Who are they kidding?’

[13] This example is cited by Prelec and Loewenstein (1998). American OnLine seems to have learned this lesson the hard way. When they offered a flat rate Internet service in early 1996 they were so overwhelmed with demand that consumers had trouble logging on to the service, causing embarrassing publicity.

Standard theory of consumer behavior does not accommodate it." Similarly, health clubs typically charge members by the month or year rather than by a per-use basis. This strategy decouples usage from fees, making the marginal cost of a visit zero. This plan is attractive because a health club is a service that many consumers feel they should use more often, but fail to do so for self-control reasons (see later). Indeed, the monthly fee, although a sunk cost, encourages use for those who want to reduce their per-visit charges. Compare this system to a pure usage-based pricing system in which Stairmaster users pay "per step." This pricing system would be completely incompatible with the psychological needs of the club member who desires usage encouragement rather than discouragement.

Perhaps the best decoupling device is the credit card. We know that credit cards facilitate spending simply by the fact that stores are willing to pay 3% or more of their revenues to the card companies (see also Feinberg 1986; Prelec and Simester 1998). A credit card decouples the purchase from the payment in several ways. First, it postpones the payment by a few weeks. This delay creates two distinct effects: (a) the payment is *later* than the purchase; (b) the payment is *separated* from the purchase. The payment delay may be attractive to some consumers who are either highly impatient or liquidity constrained, but as Prelec and Loewenstein (1998) stress, *ceteris paribus*, consumers prefer to pay before rather than after, so this factor is unlikely to be the main appeal of the credit-card purchase. Rather, the simple separation of purchase and payment appears to make the payment less salient. Along these lines, Soman (1997) finds that students leaving the campus bookstore were much more accurate in remembering the amount of their purchases if they paid by cash rather than by credit card. As he says, "Payment by credit card thus reduces the salience and vividness of the outflows, making them harder to recall than payments by cash or check, which leave a stronger memory trace" (p. 9).

A second factor contributing to the attractiveness of credit-card spending is that once the bill arrives, the purchase is mixed in with many others. Compare the impact of paying $50 in cash at the store to that of adding a $50 item to an $843 bill. Psychophysics implies that the $50 will appear larger by itself than in the context of a much larger bill, and in addition when the bill contains many items each one will lose salience. The effect becomes even stronger if the bill is not paid in full immediately. Although an unpaid balance is aversive in and of itself, it is difficult for the consumer to attribute this balance to any particular purchase.

Budgeting

So far I have been discussing mental accounting decision-making at the level of individual transactions. Another component of mental accounting is categorization or labeling. Money is commonly labeled at three levels: expenditures are grouped into budgets (e.g., food, housing, etc.); wealth is allocated into accounts (e.g., checking, pension; "rainy day"); and income is divided into categories (e.g., regular or windfall). Such accounts would be inconsequential if they were perfectly fungible (i.e., substitutable) as assumed in economics. But, they are not fungible, and so they "matter."

Consumption Categories

Dividing spending into budget categories serves two purposes. First, the budgeting process can facilitate making rational trade-offs between competing uses for funds. Second, the system can act as a self-control device. Just as organizations establish budgets to keep track of and limit divisional spending, the mental accounting system is the household's way of keeping spending within the budget (Thaler and Shefrin l98l). Of course, there is considerable variation among households in how explicit the budgeting process is.[14] As a rule, the tighter the budget, the more explicit are the budgeting rules, both in households and organizations. Families living near the poverty level use strict, explicit budgets; in wealthy families budgets are both less binding and less well defined.[15] Poorer families also tend to have budgets defined over shorter periods (a week or month), whereas wealthier families may use annual budgets. For example, Heath and Soll (1996) report that most of their MBA student subjects had weekly food and entertainment budgets and monthly clothing budgets. It is likely that these rules changed dramatically when the students got jobs at the end of their studies (in violation of the life-cycle hypothesis—see later).

Heath and Soll describe the process by which expenses are tracked against these budgets. They divide the tracking process into two stages:

> Expenses must first be noticed and [second] then assigned to their proper accounts. An expense will not affect a budget if either stage fails. To label these stages we borrow terminology from financial accounting in which the accounting system is also divided into two stages. Expenses must be booked (i.e., recorded in the accounting system) and posted (i.e., assigned to a specific expense account). Each process depends on a different cognitive system. Booking depends on attention and memory. Posting depends on similarity judgments and categorization (p. 42).[16]

Many small, routine expenses are not booked. Examples would include lunch or coffee at the workplace cafeteria (unless the norm is to bring these items from home, in which case buying the lunch might be booked). Ignoring such items is equivalent to the organizational practice of assigning small expenditures to a "petty cash" fund, not subject to the usual accounting scrutiny. The tendency to ignore small items may also explain an apparent contradiction of hedonic framing.

[14] Many of the generalizations here are based on a series of interviews conducted on my behalf in the early 1980s. See also Zelizer (1994) and her references. At one time many households used a very explicit system with envelopes of cash labeled with various spending categories. To some extent, programs such as Quicken serve as a modern replacement for this method.

[15] Still, budgets can matter even in well-off families. As the discussion of "decoupling" will later illustrate, spending on vacations may depend on whether a family rents or owns a vacation home.

[16] Regarding the categorization process, see Henderson and Peterson (1992). It should be noted that in a financial accounting system in a firm any expense that is booked is also posted.

As noted by John Gourville (1998), in many situations sellers and fund raisers elect to frame an annual fee as "pennies-a-day." Thus a $100 membership to the local public radio station might be described as a "mere 27 cents a day." Given the convex shape of the loss function, why should this strategy be effective? One possibility is that 27 cents is clearly in the petty cash category, so when the expense is framed this way it tends to be compared to other items that are not booked. In contrast, a $100 membership is large enough that it will surely be booked and posted, possibly running into binding budget constraints in the charitable-giving category. The same idea works in the opposite direction. A firm that markets a drug to help people quit smoking urges smokers to aggregate their annual smoking expenditures and think of the vacation they could take with these funds. Again, $2 a day might be ignored, but $730 pays for a nice getaway.

Implications of Violations of Fungibility

Whenever budgets are not fungible, their existence can influence consumption in various ways. One example is the case in which one budget has been spent up to its limit while other accounts have unspent funds remaining. (This situation is common in organizations. It can create extreme distortions especially if funds cannot be carried over from one year to the next. In this case one department can be severely constrained while another is desperately looking for ways to spend down this year's budget to make sure next year's is not cut.) Heath and Soll (1996) provide several experiments to illustrate this effect. In a typical study two groups of subjects were asked whether they would be willing to buy a ticket to a play. One group was told that they had spent $50 earlier in the week going to a basketball game (same budget); the other group was told that they had received a $50 parking ticket (different budget) earlier in the week. Those who had already gone to the basketball game were significantly less likely to go to the play than those who had gotten the parking ticket.[17]

Using the same logic that implies that money should be fungible (i.e., that money in one account will spend just as well in another), economists have argued that time should also be fungible. A rational person should allocate time optimally, which implies "equating at the margin." In this case, the marginal value of an extra minute devoted to any activity should be equal.[18] The jacket and calculator problem reveals that this rule does not describe choices about time. Subjects are willing to spend 20 minutes to save $5 on a small purchase but not a large one. Leclerc et al. (1995) extend this notion by reversing the problem. They ask people how much they would be willing to pay to avoid waiting in a ticket line for

[17] One might think this result could be attributed to satiation (one night out is enough in a week). However, another group was asked their willingness to buy the theater ticket after going to the basketball game for free, and they showed no effect.

[18] I am abstracting from natural discontinuities. If television shows come in increments of one hour, then one may have to choose an integer number of hours of TV watching, which alters the argument slightly.

45 minutes. They find that people are willing to pay twice as much to avoid the wait for a $45 purchase than for a $15 purchase. As in the original version of the problem, we see that the implicit value people put on their time depends on the financial context.

Self-control and Gift Giving

Another violation of fungibility introduced by the budgeting system occurs because some budgets are intentionally set 'too low' in order to help deal with particularly insidious self-control problems. For example, consider the dilemma of a couple who enjoy drinking a bottle of wine with dinner. They might decide that they can afford to spend only $10 a night on wine and so limit their purchases to wines that cost $10 a bottle on average, with no bottle costing more than $20. This policy might not be optimal in the sense that an occasional $30 bottle of champagne would be worth more than $30 to them, but they don't trust themselves to resist the temptation to increase their wine budget unreasonably if they break the $20 barrier. An implication is that this couple would greatly enjoy gifts of wine that are above their usual budget constraint. This analysis is precisely the opposite of the usual economic advice (which says that a gift in kind can be at best as good as a gift of cash, and then only if it were something that the recipient would have bought anyway). Instead the mental accounting analysis suggests that the best gifts are somewhat more luxurious than the recipient normally buys, consistent with the conventional advice (of noneconomists), which is to buy people something they wouldn't buy for themselves.

The idea that luxurious gifts can be better than cash is well known to those who design sales compensation schemes. When sales contests are run, the prize is typically a trip or luxury durable rather than cash. Perhaps the most vivid example of this practice is the experience of the National Football League in getting players to show up at the annual Pro Bowl. This all-star game is held the week after the Super Bowl and for years the league had trouble getting all of the superstar players to come. Monetary incentives were little inducement to players with seven-figure salaries. This problem was largely solved by moving the game to Hawaii and including *two* first-class tickets (one for the player's wife or girlfriend) and accommodations for all the players.

The analysis of gift giving illustrates how self-control problems can influence choices. Because expensive bottles of wine are "tempting," the couple rules them "off limits" to help control spending. For other tempting products, consumers may regulate their consumption in part by buying small quantities at a time, thus keeping inventories low. This practice creates the odd situation wherein consumers may be willing to pay a premium for a smaller quantity. This behavior is studied by Wertenbroch (1996), who finds that the price premium for sinful products in small packages is greater than for more mundane goods. His one-sentence abstract succinctly sums up his paper: "To control their consumption, consumers pay more for less of what they like too much."

Wealth Accounts

Another way of dealing with self-control problems is to place funds in accounts that are off-limits. Hersh Shefrin and I have proposed (Shefrin and Thaler 1988) that there is a hierarchy of money locations arranged by how tempting it is for a household to spend the money in each. The most tempting class of accounts is in the "current assets" category, for example cash on hand and money market or checking accounts. Money in these accounts is routinely spent each period. Less tempting to spend is money in the "current wealth" category, which includes a range of liquid asset accounts such as savings accounts, stocks and bonds, mutual funds, and so on. These funds are typically designated for saving. Next in the hierarchy is home equity. Even though the advent of home equity loans has made this category of funds somewhat less sacred, still most households aim to pay off their mortgage by the time they retire (and most succeed). Finally, in the least tempting category of funds lies the 'future income' account. These funds include money that will be earned later in life (i.e., human capital) and designated retirement savings accounts such as IRAs and 401(k)s. According to our analysis, the marginal propensity to spend a dollar of wealth in the current income account is nearly 1.0, whereas the propensity to spend a dollar of future income wealth is close to zero.

These predictions are in sharp contrast to standard economic theory of saving: the life-cycle model (Modigliani and Brumberg 1954; Friedman 1957). Here is a simplified version that captures the spirit of the life-cycle model. Suppose a person has a certain remaining lifetime of N years, and that the rate of interest is zero. Let W be the person's wealth, equal to the sum of her assets, this year's income, and future (expected) income over the rest of her life. Consumption in this period is then equal to W/N.[19] Notice that in this model any change in wealth, ΔW, no matter what form it takes (e.g., a bonus at work, an increase in the value of one's home, even an inheritance expected in a decade), produces the same change in current consumption namely $\Delta W/N$. In other words, the theory assumes that wealth is perfectly fungible.

Shefrin and I proposed a modified version of the life-cycle model, the behavioral life-cycle model, that incorporates the mental accounting temptation hierarchy described above. A powerful prediction of the mental accounting model is that if funds can be transferred to less tempting mental accounts they are more likely to be saved. This insight can be used in designing government programs to stimulate saving. According to the behavioral life-cycle model, if households can be persuaded to move some of their funds from the current income account to future income accounts, long-term savings will increase. In other words, IRAs

[19] More generally, in a world with uncertainty and positive interest rates, the life-cycle theory says that a person will spend the annuity value of his wealth in any period, that is, if he used W to buy a level annuity that paid y in every period, he would set consumption equal to y. Bequests can also be accommodated.

and 401(k)s are good vehicles to promote savings.[20] My reading of the literature on this topic is that this prediction is borne out. Households that contribute to retirement savings plans display steady increases in the funds in these accounts with no apparent reduction in the funds in other accounts. That is, they save more.[21]

Income Accounting

So far we have considered violations of fungibility produced either by the budgeting process or by the location of funds. A third class of violations can be produced by the source of the income. O'Curry (1997) investigates this phenomenon. She first has one group of subjects judge both sources and uses of funds on a serious–frivolous scale: the winnings of an office football pool are considered frivolous whereas an income tax refund is serious; eating out is frivolous but paying the bills is serious. She then asks other subjects to say what they would do with a particular windfall, such as $30 found in the pocket of a jacket in the back of the closet. She finds that people have a tendency to match the seriousness of the source of some windfall with the use to which it is put. Another example of income nonfungibility is provided by Kooreman (1997). He studies the spending behavior of families that receive child allowance payments from the Dutch government. He finds that spending on children's clothing is much more sensitive to changes in the designated child allowance than to other income sources.[22]

In the previous example the fact that the child allowance was labeled as such seemed to matter in the way people spent the money. Labeling effects are common. One surprising domain in which this idea can be applied is dividend payments by corporations. Suppose a corporation is earning profits and wishes to return some of these profits to its shareholders. One (traditional) method is to pay a dividend. Another method is simply to repurchase shares. In a world with no taxes, these two methods are equivalent. But, if (as in the United States) dividends are taxed at a higher rate than capital gains, then tax-paying shareholders would prefer share repurchases to dividends (and those who have their shares in nontaxable accounts are indifferent). Under these conditions no firm should ever pay a dividend.

[20] These accounts are especially good because not only are they less tempting 'mental' accounts but they also have a penalty for withdrawal that provides an additional incentive to leave the money in these accounts alone.

[21] See Poterba, Venti, and Wise (1996) for a current summary of the evidence supporting my claim. Their results are hotly disputed by Engen, Gale, and Scholz (1996). One reason I side with the first set of authors (aside from the fact that their results support mental accounting) is that the simplest analyses show that the savings plans increase saving. Obtaining the opposite results seems to require a lot more work.

[22] There is a similar finding in public finance called the "flypaper effect." When local governments receive earmarked payments for particular kinds of expenditure (e.g., schools) they tend to increase their spending on that activity by the full amount of the grant. Economic theory predicts that they would increase their spending only by the fraction of their income that they normally spend on this activity. See Hines and Thaler (1995).

Why do firms pay dividends? Shefrin and Statman (1984) have proposed an explanation based on mental accounting. They argue that investors like dividends because the regular cash payment provides a simple self-control rule: spend the dividends and leave the principal alone. In this way, the dividend acts like an allowance. If, instead, firms simply repurchased their own shares, stockholders would not receive a designated amount to spend, and would have to dip into capital on a period basis. Retirees (who tend to own high-dividend-paying stocks) might then worry that they would spend down the principal too quickly. A similar nonfungibility result is offered by Hatsopoulos, Krugman, and Poterba (1989). Although capital gains in the stock market tend to have little effect on consumption, these authors found that when takeovers generate cash to the stockholders, consumption does increase. This is sometimes called the "mailbox effect." When the check arrives in the mailbox it tends to get spent. Gains on paper are left alone.

Choice Bracketing and Dynamic Mental Accounting

A recurring theme of this chapter is that choices are altered by the introduction of notional (but nonfungible) boundaries. The location of the parentheses matters in mental accounting—a loss hurts less if it can be combined with a larger gain; a purchase is more likely to be made if it can be assigned to an account that is not already in the red; and a prior (sunk) cost is attended to if the current decision is in the same account. This section elaborates on this theme by considering other ways in which boundaries are set, namely whether a series of decisions are made one at a time or grouped together (or "bracketed," to use the language of Read, Loewenstein, and Rabin 1998).

Prior Outcomes and Risky Choice

In their prospect paper, Kahneman and Tversky mention the empirical finding that betting on long shots increases on the last race of the day, when the average bettor is (i) losing money on the day, and (ii) anxious to break even.[23] An interesting feature of this sunk cost effect is that it depends completely on the decision to close the betting account daily. If each race were a separate account, prior races would have no effect, and similarly if today's betting were combined with the rest of the bettor's wealth (or even his lifetime of bets), the prior outcome would likely be trivial.

This analysis applies to other gambling decisions. If a series of gambles are bracketed together, then the outcome of one gamble can affect the choices made later. Johnson and I investigated how prior outcomes affect risky choice (Thaler and Johnson 1990). Subjects were MBA students who played for real money. The

[23] That is, long shots become even worse bets at the end of the day. They are always bad bets. See Thaler and Ziemba (1988).

following three choices illustrate the type of problems studied. The percentage of subjects taking each option appears in brackets.

Problem 1. You have just won $30. Now choose between:
(a) A 50% chance to gain $9 and a 50% chance to loose $9. [70]
(b) No further gain or loss. [30]

Problem 2. You have just lost $30. Now choose between:
(a) A 50% chance to gain $9 and a 50% chance to loose $9. [40]
(b) No further gain or loss. [60]

Problem 3. You have just lost $30. Now choose between:
(a) A 33% chance to gain $30 and a 67% chance to gain nothing. [60]
(b) A sure $10. [40]

These and other problems of this sort were used to investigate how prior outcomes affect risky choices. Two results are worth noting. First, as illustrated by Problem 1, a prior gain can stimulate risk seeking in the same account. We called this phenomenon the 'house money' effect since gamblers often refer to money they have won from the casino as house money (the casino is known as 'the house'). Indeed, one often sees gamblers who have won some money early in the evening put that money into a different pocket from their 'own' money; this way each pocket is a separate mental account. Second, as illustrated by Problems 2 and 3, prior losses did not stimulate risk seeking unless the gamble offered a chance to break even.

The stakes used in the experiments just described were fairly large in comparison to most laboratory experiments, but small compared to the wealth of the participants. Limited experimental budgets are a fact of life. Gertner (1993) has made clever use of a set of bigger stakes choices over gambles made by contestants on a television game show called "Card Sharks."[24] The choices Gertner studies were the last in a series of bets made by the winner of the show that day. The contestant had to predict whether a card picked at random from a deck would be higher or lower than a card that was showing. Aces are high and ties create no gain or loss. The odds on the bet therefore vary from no risk (when the showing card is a 2 or an Ace) to roughly 50–50 when the up-card is an 8. After making the prediction, the contestant then can make a bet on the outcome, but the bet must be between 50% and 100% of the amount she has won on the day's show (on average, about $3000). Ignoring the sure bets, Gertner estimates a Tobit regression model to predict the size of the contestant's bet as a function of the card showing (the odds), the stake available (that is, today's winnings), and the amount won in previous days on the show. After controlling for the constraint that the bet must lie between 50% and 100% of the stake, Gertner finds that today's winnings strongly

[24] See also Biswanger (1981), who obtains similar results. He also was able to run high stakes experiments by using subjects in rural villages in India.

influences on the amount wagered.[25] In contrast, prior cash won has virtually no effect. This finding implies that cash won *today* is treated in a different mental account from cash won the day before.[26] This behavior is inconsistent with any version of expected utility theory that treats wealth as fungible.

Narrow Framing and Myopic Loss-Aversion

In the gambling decisions discussed above, the day of the experiment suggested a natural bracket. Often gambles or investments occur over a period of time, giving the decision-maker considerable flexibility in how often to calculate gains and losses. It will come as no surprise to learn that the choice of how to bracket the gambles influences the attractiveness of the individual bets. An illustration is provided by a famous problem first posed by Paul Samuelson. Samuelson, it seems, was having lunch with an economist colleague and offered his colleague an attractive bet. They would flip a coin, and if the colleague won he would get \$200; if he lost he would have to pay only \$100. The colleague turned this bet down, but said that if Samuelson would be willing to play the bet 100 times he would be game. Samuelson (1963) declined to offer this parlay, but went home and proved that this pair of choices is irrational.[27]

There are several points of interest in this problem. First, Samuelson quotes his colleague's reasoning for rejecting the single play of the gamble: "I won't bet because I would feel the \$100 loss more than the \$200 gain." Modern translation "I am loss-averse." Second, why does he like the series of bets? Specifically, what mental accounting operation can he be using to make the series of bets attractive when the single play is not?

Suppose Samuelson's colleague's preferences are a piecewise linear version of the prospect theory value function with a loss-aversion factor of 2.5:

$$U(x) = \begin{cases} x & x \geqslant 0 \\ 2.5x & x < 0 \end{cases}$$

Because the loss-aversion coefficient is greater than 2, a single play of Samuelson's bet is obviously unattractive. What about two plays? The attractiveness of two bets depends on the mental accounting rules being used. If each play of the bet is treated as a separate event, then two plays of the gamble are twice as bad as one play. However, if the bets are combined into a portfolio, then the two-bet

[25] Gertner offers the following example to illustrate this difference. Suppose a first-time contestant has won \$5000 so far and has a Jack showing, so a bet of 'lower' offers 3–1 odds. (She loses with an A, K, or Q, ties with a J, and wins otherwise.) The regression predicts a bet of \$2800. Compare this contestant to one who has won only \$3000 today but won \$2000 the previous day. Although their winnings on the show are identical, this player is predicted to bet only \$1544.

[26] This result is all the more striking because 'yesterday's' show was probably taped just an hour before 'today's' (several shows are taped in the same day) and 'yesterday's' winnings have certainly not been collected.

[27] Specifically, he showed that an expected utility maximizer who will not accept a single play of a gamble for any wealth level that could obtain over a series of such bets will not accept the series. For a more general result, see Tversky and Bar Hillel (1983).

parlay {\$400, 0.25; 100, 0.50; −\$200, 0.25} yields positive expected utility with the hypothesized utility function, and as the number of repetitions increases the portfolio becomes even more attractive. So Samuelson's colleague should accept any number of trials of this bet strictly greater than one *as long as he does not have to watch*!

More generally, loss-averse people are more willing to take risks if they combine many bets together than if they consider them one at a time. Indeed, although the puzzle to Samuelson was why his colleague was willing to accept the series of bets, the real puzzle is why he was unwilling to play one. Risk-aversion cannot be a satisfactory explanation if his colleague has any significant wealth. For example, suppose Samuelson's colleague's utility function is $U(W) = \ln W$ and his wealth is a modest \$10,000. In that case he should be willing to risk a 50% chance of losing \$100 if he had a 50% chance to gain a mere \$101.01! Similar results obtain for other reasonable utility functions. In fact, Rabin (1998) shows that expected utility theory implies that someone who turns down Samuelson's bet should also turn down a 50% chance to lose \$200 and a 50% chance to win \$20,000. More generally, he shows that expected-utility theory requires people to be virtually risk neutral for "small" bets. To explain the fact that many people do reject attractive small bets (such as Samuelson's), we need a combination of loss aversion and one-bet-at-a-time mental accounting.

Benartzi and I (1995) use the same analysis to offer a mental accounting explanation for what economists call the equity premium puzzle (Mehra and Prescott 1985). The equity premium is the difference in the rate of return on equities (stocks) and a safe investment such as treasury bills. The puzzle is that this difference has historically been very large. In the United States the equity premium has been roughly 6% per year over the past 70 years. This means that a dollar invested in stocks on 1 January 1926 was worth more than \$1800 on 1 January 1998, whereas a dollar invested in treasury bills was worth only about \$15 (half of which was eaten up by inflation). Of course, part of this difference can be attributed to risk, but what Mehra and Prescott show is that the level of risk aversion necessary to explain such a large difference in returns is implausible.[28]

To explain the puzzle we note that the risk attitude of loss-averse investors depends on the frequency with which they reset their reference point, i.e. how often they 'count their money'. We hypothesize that investors have prospect theory preferences (using parameters estimated by Tversky and Kahneman 1992).[29] We then ask how often people would have to evaluate the changes in their portfolios to make them indifferent between the (US) historical distributions of returns on stocks and bonds? The results of our simulations suggest that the answer is about 13 months. This outcome implies that if the most prominent evaluation period for investors is once a year, the equity premium puzzle is "solved."

[28] They estimate that it would take a coefficient of relative risk-aversion of about 40 to explain the history equity premium. In contrast, a log utility function has a coefficient of 1.

[29] Specifically, the value function is: $v(x) = x^{\alpha}$ if $x \geq 0$ $-\lambda(-x)^{\beta}$ if $x < 0$ where λ is the coefficient of loss-aversion. They have estimated α and β to be 0.88 and λ to be 2.25. We also use their rank-dependent weighting function. For details see Benartzi and Thaler (1995).

We refer to this behavior as myopic loss-aversion. The disparaging term "myopic" seems appropriate because the frequent evaluations prevent the investor from adopting a strategy that would be preferred over an appropriately long time-horizon. Indeed, experimental evidence supports the view that when a long-term horizon is imposed externally, subjects elect more risk. For example, Gneezy and Potters (1997) and Thaler et al. (1997) ran experiments in which subjects make choices between gambles (investments). The manipulations in these experiments are the frequency with which subjects get feedback. For example, in the Thaler et al. study, subjects made investment decisions between stocks and bonds at frequencies that simulated either eight times a year, once a year, or once every five years. The subjects in the two long-term conditions invested roughly two-thirds of their funds in stocks while those in the frequent evaluation condition invested 59% of their assets in bonds. Similarly, Benartzi and I (forthcoming) asked staff members at a university how they would invest their retirement money if they had to choose between two investment funds, A and B, one of which was based on stock returns, the other on bonds. In this case the manipulation was the way in which the returns were displayed. One group examined a chart showing the distribution of *one-year* rates of return, and the other group was shown the simulated distribution of 30-*year* rates of return. Those who saw the one-year returns said they would invest a majority of their funds in bonds, whereas those shown the 30-year returns invested 90% of their funds in stocks.[30]

Myopic loss-aversion is an example of a more general phenomenon that Kahneman and Lovallo (1993) call narrow framing; projects are evaluated one at a time, rather than as part of an overall portfolio. This tendency can lead to an extreme unwillingness to take risks. I observed an interesting illustration of this phenomenon while teaching a group of executives from one firm, each of whom was responsible for managing a separate division. I asked each whether he would be willing to undertake a project for his division if the payoffs were as follows: 50% chance to gain $2 million, 50% chance to lose $1 million. Of the 25 executives, three accepted the gamble. I then asked the CEO, who was also attending the session, how he would like a portfolio of 25 of these investments. He nodded enthusiastically. This story illustrates that the antidote for excessive risk aversion is aggregation, either across time or across different divisions.

The examples discussed so far show that narrow bracketing can inhibit risk-taking. Narrow bracketing can also have other perverse side-effects. For example, Camerer et al. (1997) study the daily labor supply decisions of New York City taxi drivers. In New York, as in many cities, the cab drivers typically rent their cars for a 12-hour period for a fixed fee. They are then entitled to keep all the revenues they earn during that half-day. Since 12 hours is a long time to drive a car, especially in New York City, the drivers must decide each day how long to drive; that is, whether to keep the car for the full 12 hours or quit earlier. This decision is complicated by the fact that there is more demand for their services on some days

[30] Similar results for gambles are also obtained by Keren and Wagenaar (1987) and Redelmeier and Tversky (1992).

than others (because of differences in weather or the presence of a big convention, for example). A rational analysis would lead drivers to work longer hours on busy days, as this policy would maximize earnings per hour worked. If, instead, drivers establish a target earnings level *per day*, they will tend to quit earlier on good days. This is precisely what Camerer et al. find. The elasticity of hours worked with respect to the daily wage (as measured by the earnings of *other drivers that day*) is strongly negative. The implication is that taxi drivers do their mental accounting one day at a time.[31]

The Diversification Heuristic

The unit of analysis can also influence how much variety consumers elect. This effect was first demonstrated by Simonson (1990). He gave students the opportunity to select among six snacks (candy bars, chips, etc.) in one of two conditions: (a) sequential choice: they picked one of the six snacks at each of three class meetings held a week apart; (b) simultaneous choice: on the first class meeting they selected three snacks to be consumed one snack per week over the three class meetings. Simonson observed that in the simultaneous choice condition subjects displayed much more variety seeking than in the sequential choice condition. For example, in the simultaneous choice condition 64% of the subjects chose three different snacks whereas in the sequential choice condition only 9% of the subjects made this choice. Simonson suggests that this behavior might be explained by variety seeking serving as a choice heuristic. That is, when asked to make several choices at once, people tend to diversify. This strategy is sensible under some circumstances (such as when eating a meal—we typically do not order three courses of the same food), but can be misapplied to other situations, such as sequential choice. This mistake represents a failure of predicted utility to accurately forecast subsequent experienced utility. Many students who liked Snickers best elected that snack each week when they picked one week at a time, but went for variety when they had to choose in advance.

This result has been called the "diversification bias" by Read and Loewenstein (1995). They demonstrate the role of choice bracketing in an ingenious experiment conducted on Halloween night. The "subjects" in the experiment were young trick-or-treaters who approached two adjacent houses. In one condition the children were offered a choice between two candies (Three Musketeers and Milky Way) at each house. In the other condition they were told at the first house they reached to "choose whichever two candy bars you like." Large piles of both candies were displayed to assure that the children would not think it rude to take two of the same. The results showed a strong diversification bias in the simultaneous choice condition: every child selected one of each candy. In contrast, only 48% of the children in the sequential choice condition picked different candies. This result is striking, since in either case the candies are dumped into a bag and

[31] Rizzo and Zeckhauser (1998) find a similar result for physicians whose evaluation period appears to be one year rather than one day.

consumed later. It is the portfolio in the bag that matters, not the portfolio selected at each house.

The diversification bias is not limited to young people choosing among snacks. Benartzi and I (1998) have found evidence of the same phenomenon by studying how people allocate their retirement funds across various investment vehicles. In particular, we find some evidence for an extreme version of this bias that we call the 1/*n* heuristic. The idea is that when an employee is offered *n* funds to choose from in her retirement plan, she divides the money evenly among the funds offered. Use of this heuristic, or others only slightly more sophisticated, implies that the asset allocation an investor chooses will depend strongly on the array of funds offered in the retirement plan. Thus, in a plan that offered one stock fund and one bond fund, the average allocation would be 50% stocks, but if another stock fund were added, the allocation to stocks would jump to two thirds. We find evidence supporting just this behavior. In a sample of pension plans we regress the percentage of the plan assets in stocks on the percentage of the funds that are stock funds and find a very strong relationship.

We also find that employees seem to put stock in the company they work for into a separate mental account. For companies that do not offer their own stock as one of the options in the pension plan the employees invest 49% of their money in bonds and 51% in stocks. When the company stock is included in the plan this investment attracts 42% of the funds. If the employees wanted to attain a 50% equity exposure, they would invest about 8% of the rest of their funds in stocks, the rest in bonds. Instead they invest their non-company stock funds evenly: 29% in stocks, 29% in bonds.

Discussion

My own thinking about mental accounting began with an attempt to understand why people pay attention to sunk costs, why people are lured by bargains into silly expenditures, and why people will drive across town to save $5 on a small purchase but not a large one. I hope this paper has shown that we have learned quite a bit about these questions, and in so doing, the researchers working in this area have extended the scope of mental accounting far beyond the original set of questions I had set out to answer. Consider the range of questions that mental accounting helps us answer:

- Why do firms pay dividends?
- Why do people buy time-share vacation properties?
- Why are flat-rate pricing plans so popular?
- Why do sales contests have luxuries (instead of cash) as prizes?
- Why do 401(k) plans increase savings?
- Why do stocks earn so much higher a return than bonds?
- Why do people decline small-stakes attractive bets?
- Why can't you get a cab on a rainy day? (Hint: cab drivers earn more per hour on rainy days.)

A question that has not received much attention is whether mental accounting is good for us. What is the normative status of mental accounting? I see no useful purpose in worrying about whether or not mental accounting is 'rational'. Mental accounting procedures have evolved to economize on time and thinking costs and also to deal with self-control problems. As is to be expected, the procedures do not work perfectly. People pay attention to sunk costs. They buy things they don't need because the deal is too good to pass up. They quit early on a good day. They put their retirement money in a money market account.

It is not possible to say that the system is flawed without knowing how to fix it. Given that optimization is not feasible (too costly) repairing one problem may create another. For example, if we teach people to ignore sunk costs, do they stop abiding by the principle "waste not, want not"? If we stop being lured by good deals, do we stop paying attention to price altogether? There are no easy answers.

Those interested in improving individual decision making can do more work on mental accounting as a prescriptive device. How can mental-accounting rules be modified to achieve certain goals?[32] For example, Jonathan Clements, the author of a regular column for new investors in the *Wall Street Journal* called "Getting Going" invited readers to submit tips on how to do a better job of saving and investing.[33] Many of the tips he later published had a strong mental accounting flavor. One reader, David Guerini, submitted the following advice:

> I started a little "side" savings account eight years ago. During the day, I try to accumulate change. If I spend $4.50 at a store, I give the cashier a $5 bill, even if I have 50 cents in my pocket. At the end of each day, the money is put aside. If I have no change, I put a $1 bill aside. I add income-tax refunds, money from products I purchased and returned for a refund, and all those annoying little mail-in rebates they give you when you purchase batteries, shaving cream, and so on. I end up painlessly saving between $500 and $1000 each year.

An economist might argue that it would be even less painful just to write a check once a year and send it to his mutual fund. But that would miss the point: mental accounting matters.

References

Ainslie, George. 1975. "Special Reward: A Behavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin*, 82(4): 463–96.

Arkes, Hal R., and Catherine Blumer. "The Psychology of Sunk Cost." *Organizational Behavior and Human Decision Processes*. 35(1): 124–40.

Benartzl, Shlemo, and Richard H. Thaler. 1995. "Myopic Loss-Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics*, 110: 75–92.

[32] Along these lines, Read, Loewenstein and Rabin (1998) have a usefuo discussion of when broad bracketing works better than narrow bracketing. Short answer usually

[33] See his column on 20, 24, and 31 January 1998.

———. In press. “Risk Aversion or Myopia: Choices in Repeated Gambles and Retirement Investments,” *Management Science*.

———. 1998. “Illusory Diversification and Retirement Savings.” Working paper, University of Chicago and UCLA.

Biswanger, Hans. 1981. “Attitudes toward Risk: Theoretical Implications of an Experiment in Rural India.” *Economic Journal*, 91: 867–90.

Burgstahler, Dave, and Ilic Dichev. 1997. Earnings Management to Avoid Earnings Decreases and Losses.” *Journal of Accounting and Economics*, 24: 99–126.

Camerer, C., Linda Babcock, George Loewenstein, and Richard H. Thaler. 1997. “Labor Supply of New York City Cabdrivers: One Day at a Time,” *Quarterly Journal of Economics*, 112: 407–442.

Degeorge, Francois, Jay Patel, and Richard J. Zeckhauser. In press. “Earnings Management to Exceed Thresholds.” *Journal of Business*.

Engen, Eric., William Gale, and John K. Scholz. 1996. “The Illusory Effects of Saving Incentive Programs,” *Journal of Economics Perspectives*, 10: 113–38.

Feenberg, D., and Jonathan Skinner. 1989. “Sources of IRA saving.” In *Tax Policy and the Economy*: *Vol. 3*, edited by L. Summers. Cambridge: MIT Press.

Feinberg, R. A. 1986. “Credit Cards as Spending Facilitating Stimuli: A Conditioning Interpretation.” *Journal of Consumer Research*, 12: 304–56.

Fishburn, Peter, and Duncan Luce. 1995. Joint Receipt and Thaler’s Hedonic Editing Rule.” *Mathematical Social Science*, 29: 33–76.

Friedman, Milton. 1957. *A Theory of Consumption Function*. Princeton: Princeton University Press.

Gertner, Robert. 1993. “Game Shows and Economic Behavior: Risk Taking on ‘Card Sharks.’ ” *Quarterly Journal of Economics*, 106: 507–21.

Gneezy, Uri, and Jan Potters. 1997. An Experiment on Risk Taking and Evaluating Periods.” *Quarterly Journal of Economics*, 112(May): 631–46.

Gourville, John. 1998. “Pennies-a-Day: The Effect of Temporal Re-framing on Transaction Evaluation.” *Journal of Consumer Research*, 24: 395–408.

Gourville, John, and Dilip Soman. 1998. “Payment Depreciation: The Effects of Temporally Separating Payments from Consumption.” *Journal of Consumer Research*.

Hatsopoulos, George N., Paul R. Krugman, and James M. Poterba. 1989. “Overconsumption: The Challenge to U.S. Economic Policy.” *American Business Conference*.

Heath, Chip. 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.

Heath, Chip, and Jack B. Soll. 1996. “Mental Accounting and Consumer Decisions.” *Journal of Consumer Research*, 23: 40–52.

Henderson, Pamela, and Robert Peterson. 1992. “Mental Accounting and Categorization.” *Organizational Behavioral and Human Decision Processes*, 51: 92–117.

Hines, James, and Richard H. Thaler. 1995. “Anomalies: The Flypaper Effect.” *Journal of Economics Perspectives*, 9: 217–26.

Kahneman, Daniel. 1992. “Reference Points, Anchors, Norms, and Mixed Feelings.” *Organizational Behavior and Human Decision Processes*, 51: 296–312.

———. 1994. “New Challenges to the Rationality Assumption.” *Journal of Institutional and Theoretical Economics*, 150: 43–56.

Kahneman, Daniel, Jack Knetsch, and Richard H. Thaler. 1986. “Fairness as a Constraint on Profit-Seeking: Entitlements in the Market.” *American Economic Review*, 76: 728–41.

Kahneman, Daniel, and Dan Lovallo. 1993. "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking." *Management Science*, 39: 17–31.

Kahneman, Daniel, and Ames Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47: 263–91.

———. 1984. "Choices, Values, and Frames." The American Psychologist, 39: 341–50.

Keren, Gideon. 1991. "Additional Tests of Utility Theory under Unique and Repeated Conditions." *Journal of Behavioral Decision Making*, 4: 297–304.

Keren, Gideon, and Willem A. Wagenaar. 1987. "Violation of Utility Theory in Unique and Repeated Gambles." *Journal of Experiemental Psychology: Learning, Memory and Cognition*, 13: 387–91.

Kooreman, P. 1997. "The Labeling Effect of a Child Benefit System." Unpublished working paper, University of Groningen.

Leclerc, France, Bernd Schmidt, and Laurette Dube. 1995. "Decision Making and Waiting Time: Is Time like Money?" *Journal of Consumer Research*, 22: 110–19.

Linville, Patricia, and Gref W. Fischer. 1991. "Preferences for Separating or Combining Events." *Journal of Personality and Social Psychology*, 60: 5–23.

Mehra, Rajiv, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics*, 15: 145–61.

Modigliani, France, and R. Brumberg. 1954. "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data." In *Post Keynesian Economics*, edited by K. K. Kurihara. New Brunswick, NJ: Rutgers University Press.

O'Curry, Susan. 1997. "Income Source Effects." Unpublished working paper, DePaul University.

Odean, Terrance. 1998. "Are Investors Reluctant to Realize Their Losses." *Journal of Finance* (October).

Poterba, James M., Steve F. Ventl, and David A. Wise. 1996. "How Retirement Savings Programs Increase Saving." *Journal of Economics Perspectives*, 10(Fall): 91–112.

Prelec, Drazen, and George Loewenstein. "The Red and the Black: Mental Accounting of Savings and Debt." *Marketing Science*, 17: 4–28.

Prelec, Drazen, and Duncan Simester. 1998. "Always Leave Home without It." Working paper, MIT Sloan School.

Rabin, Matthew. 1998. "Risk Aversion, Diminishing Marginal Utility, and Expected-Utility Theory: A Calibration Theorem." Working paper, Department of Economics, University of California, Berkeley.

Read, Daniel, and George Loewenstein. 1995. "Diversification Bias: Explaining the Discrepancy in Variety Seeking between Combined and Separated Choices." *Journal of Experimental Psychology: Applied*, 1: 34–49.

Read, Daniel, George Loewenstein, and Matthew Rabin. 1998. "Choice Bracketing." Unpublished working paper, Carnegie Mellon University, 1998.

Redelmeier, Don A., and Amer Tversky. 1992. "On the Farming of Multiple Prospects." *Psychological Science*, 3(3): 191–93.

Rizzo, J. A., and Richard J. Zeckhauser. 1998. "Income Targets and Physician Behavior." Working paper, JFK School of Government, Harvard University.

Samuelson, Paul A. 1963. "Risk and Uncertainty: A Fallacy of Large Numbers." *Scientia*, 108–13.

Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: John Wiley.

Shafir, Elder, and Richard H. Thaler. 1998. "Invest Now, Drink Later, Spend Never: The Mental Accounting of Advanced Purchases." Working paper, Graduate School of Business University of Chicago.

Shefrin, Hersh M., and Meier Statman. 1984. “Explaining Investor Preference for Cash Dividends.” *Journal of Financial Economics*, 13: 253–82.

———. 1987. “The Disposition to Sell Winners too Early and Ride Losers too Long.” *Journal of Finance*, 40: 777–90.

Shefrin, Hersh M., and Richard H. Thaler. 1988. “The Behavioral Life-Cycle Hypothesis.” *Economic Inquiry*, 26(October): 609–43.

Simonson, Itamar. 1990. “The effect of Purchase Quantity and Timing on Variety-Seeking Behavior.” *Journal of Marketing Research*, 28: 150–62.

Soman, Dilip. 1997. “Contextual Effects of Payment Mechanism on Purchase Intention: Check or Charge?” Unpublished working paper, University of Colorado.

Thaler, Richard H. 1980. “Toward a Positive Theory of Consumer Choice.” *Journal of Economic Behavior and Organization*, 1: 39–60.

———. 1985. Mental Accounting and Consumer Choice.” *Marketing Science*, 4: 199–214.

———. 1990. “Saving, Fungibility and Mental Accounts.” *Journal of Economic Perspectives*, 4: 193–205.

Thaler, Richard H., and Eric J. Johnson. 1990. “Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice.” *Management Science*, 36(6): 643–60.

Thaler, Richard H., and Hersh M. Shefrin. 1981. “An Economic Theory of Self-Control.” *Journal of Political Economy*, 39: 392–406.

Thaler, Richard H., Amos Tversky, Daniel Kahneman, and Alan Schwartz. 1997. “The Effect of Myopia and Loss Aversion on Risk Taking: An Experiemental Test.” *Quarterly Journal of Economics* 112: 647–61.

Thaler, Richard H., and William Ziemba. 1988. “Parl-mutual Betting Markets: Racetracks and Lotteries.” *Journal of Economic Perspectives*, 2: 161–74.

Train, Kenneth E. 1991. *Optimal Regulation* Cambridge, MA: MIT Press.

Tversky, Ames, and Maya Bar-Hillel. The Long Run and the Short.” *Journal of Experiemental Psychology: Human Learning Memory, and Cognition*, 9: 713–17.

Tversky, Ames, and Daniel Kahneman. 1981. “The Framing of Decisions and the Rationality of Choice.” *Science*, 211: 453–58.

——— 1992. “Advances in Prospect Theory: Cumulative Representation of Uncertainty.” *Journal of Risk and Uncertainty*, 5: 297–323.

Wertenbroch, Klaus. 1996. “Consumption Self-control via Purchase Rationing.” Working paper, Yale University.

Zelizer, Viviana. A. *The Social Meaning of Money: Pin Money, Paychecks, Poor Relief, and Other Currencies*. New York: Basic Books, 1994.

CHAPTER 4

Developments in Nonexpected-Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk

CHRIS STARMER

1. Introduction

How many theories of decision making under risk and uncertainty can you think of? Readers of this article will no doubt be familiar with expected-utility theory (EU), the standard theory of individual choice in economics. Many, I expect, will know of a few alternatives to this model. But how many, I wonder, will be aware that these so-called nonexpected utility models now number well into double figures? An enormous amount of theoretical effort has been devoted toward developing alternatives to EU, and this has run hand-in-hand with an ongoing experimental program aimed at testing those theories. The good and proper division of labor suggests that a relatively small group of specialists will be fully aware of the details of this literature. At the same time, the implications of developments in this field are of more than passing interest to the general economist since what stimulated developments in non-EU is surely of widespread concern: put bluntly, the standard theory did not fit the facts.

As the standard theory of individual decision making, and as a core component of game theory, EU constitutes a key building block of a vast range of economic theory. It should be no surprise, therefore, that developing a better understanding of the determinants of individual choice behavior seemed a natural research priority to many theorists. Around two decades of quite intensive research on the topic has generated a great deal of theoretical innovation plus a much richer body of evidence against which models can be judged. There can be few areas in economics that could claim to have sustained such a rich interaction between theory and evidence in an ongoing effort to develop theories in closer conformity with the facts. Considered together, the accumulated theory and evidence present an opportunity to reflect on what has been achieved. Perhaps the most obvious question to address to this literature is this: Has it generated, or does it show the prospect of generating, a serious contender for replacing EU, at least for certain purposes?

I owe thanks to Colin Camerer, Robin Cubitt, Paolo Ghirardato, Mark Machina, John Quiggin, Uzi Segal, Robert Sugden, Peter Wakker, and George Wu, plus an anonymous referee for extremely helpful comments on and discussions around this chapter. I am also grateful for support from the Economic and Social Research Council of the UK (Risk and Human Behavior Research Programme).

In what follows, my aim will be to set out what I take to have been key theoretical developments in the area, to review the related evidence, and to draw conclusions about the current state of play and the prospects for the future. In doing so, rather than simply present an exhaustive list of models, my aim will be to identify and discuss different modeling *strategies*, picking specific models as illustrations. I also intend to narrow my sights in two significant respects. First, my focus will be on *descriptive* as opposed to *normative* issues. Second, I will concentrate on the problem of modeling choices under *risk* as opposed to the more general category of *uncertainty* (the distinction is explained in the next section). Clearing the ground in this way will, I hope, sharpen the focus on one central research problem that continues to motivate much of the research in this arena: the endeavor to develop a "satisfactory" account of *actual* decision behavior in situations of *risk*. It will be a personal view, but one which I hope will help the interested nonspecialist find a trail through this expansive and quite detailed literature.

The chapter is organized as follows. Sections 2 and 3 set the scene with discussions of the standard theory and the evidence that prompted theorists to look for alternatives. Section 4 provides the core overview of nonexpected utility theories. Section 5 seeks to evaluate what has been achieved so far, and in three subsections I discuss (1) how new theories have fared in a second phase of experimental testing, (2) how new theories may help us to explain a range of phenomena "in the field," and (3) whether nonexpected utility theory offers a viable alternative to EU for everyday theoretical use.

2. Where It Began

Although the primary purpose of this chapter is to review alternatives to EU, that theory provides the natural point of departure since most of the theories I will be discussing can be understood as generalizations of this base theory.[1] EU was first proposed by Daniel Bernoulli (1738) in response to an apparent puzzle surrounding what price a reasonable person should be prepared to pay to enter a gamble. It was the conventional wisdom at the time that it would be reasonable to pay anything up to the expected value of a gamble, but Bernoulli presents this counterexample: A coin is flipped repeatedly until a head is produced; if you enter the game, you receive a payoff of, say, $\$2^n$ where n is the number of the throw producing the first head. This is the so-called St. Petersburg game. It is easy to see that its expected monetary payoff is infinite, yet Bernoulli believed that most people would be prepared to pay only a relatively small amount to enter it, and he took this intuition as evidence that the "value" of a gamble to an individual is not, in general, equal to its expected monetary value. He proposed a theory in which individuals place subjective values, or "utilities," on monetary outcomes and the value of a gamble is the expectation of these utilities. While Bernoulli's theory—the

[1] I shall not dwell on this account of EU. For those interested in further discussion, an excellent starting place is Paul Schoemaker's (1982) review.

first statement of EU—solved the St. Petersburg puzzle, it did not find much favor with modern economists until the 1950s. This is partly explained by the fact that, in the form presented by Bernoulli, the theory presupposes the existence of a cardinal utility scale; an assumption that did not sit well with the drive toward ordinalization during the first half of the twentieth century.

Interest in the theory was revived when John von Neuman and Oskar Morgenstern (1947) showed that the expected utility hypothesis could be derived from a set of apparently appealing axioms on preference. Since then, numerous alternative axiomatizations have been developed, some of which seem highly appealing, some might even say compelling, from a normative point of view (see for example Peter Hammond 1988).[2] To the extent that its axioms can be justified as sound principles of rational choice to which any reasonable person would subscribe, they provide grounds for interpreting EU *normatively* (as a model of how people ought to choose) and *prescriptively* (as a practical aid to choice). My concern, however, is with how people actually choose, whether or not such choices conform with a priori notions of rationality. Consequently, I will not be delayed by questions about whether particular axioms can or cannot be defended as sound principles of rational choice, and I will start from the presumption that evidence relating to actual behavior should not be discounted purely on the basis that it falls foul of conventional axioms of choice.

For the purpose of understanding alternative models of choice, it will be useful to present one set of axioms from which EU can be derived. In the approach that I adopt, at least to begin with, preferences are defined over *prospects*, where a prospect is to be understood as a list of consequences with associated probabilities. I will assume throughout that all consequences and probabilities are known to the agent, and hence, in choosing among prospects, the agent can be said to confront a situation of *risk* (in contrast to situations of *uncertainty* in which at least some of the outcomes or probabilities are unknown). I will use lowercase letters in bold (e.g., q, r, s) to represent prospects, and the letter p to represent probabilities (take it that p always lies in the interval [0,1]). A given prospect may contain other prospects as consequences, but assuming that such compound prospects can be reduced to simple prospects following the conventional rules of probability, we can represent any prospect q by a probability distribution $q = (p_1, \ldots, p_n)$ over a fixed set of pure consequences $X = (x_1, \ldots, x_n)$ where p_i is the probability of x_i, $p_i \geq 0$ for all i, and • $p_i = 1$. Hence, the elements of X are to be understood as an exhaustive and mutually exclusive list of possible consequences which may follow from a particular course of action. While this notation allows a prospect to be written simply as vector of probabilities (as q above) it will sometimes be useful to be explicit about the consequences too—e.g., by writing $q = (x_1, p_1, \ldots, x_n, p_n)$.

Given these preliminaries, the expected utility hypothesis can be derived from three axioms: ordering, continuity, and independence. The ordering axiom requires

[2] Such arguments, while widely accepted, are nevertheless controversial. See, for example, Anand (1993) and Sugden (1991).

both completeness and transitivity. Completeness entails that for all q, r: *either* $q \succcurlyeq r$ *or* $r \succcurlyeq q$ *or both* where $\succcurlyeq$ represents the relation "is (weakly) preferred to." Transitivity requires that for all q, r, s: *if* $q \succcurlyeq r$ *and* $r \succcurlyeq s$, *then* $q \succcurlyeq s$. Continuity requires that for all prospects q, r, s where $q \succcurlyeq r$ *and* $r \succcurlyeq s$, *there exists some* p *such that* $(q, p; s, 1 - p) \sim r$, where $\sim$ represents the relation of indifference and $(q, p; s, 1 - p)$ represents a (compound) prospect that results in q with probability p; s with probability $1 - p$. Together the axioms of ordering and continuity imply that preferences over prospects can be represented by a function $V(\cdot)$ which assigns a real-valued index to each prospect. The function $V(\cdot)$ is a representation of preference in the sense that $V(q) \bullet V(r) \Leftrightarrow q \succcurlyeq r$: that is, an individual will choose the prospect q over the prospect r if, and only if, the value assigned to q by $V(\cdot)$ is no less than that assigned to r.

To assume the existence of some such preference function has seemed, to many economists, the natural starting point for any economic theory of choice; it amounts to assuming that agents have well-defined preferences, while imposing minimal restriction on the precise form of those preferences. For those who endorse such an approach, the natural questions center around what further restrictions can be placed on $V(\cdot)$? The independence axiom of EU places quite strong restrictions on the precise form of preferences: it is this axiom which gives the standard theory most of its empirical content (and it is the axiom that most alternatives to EU will relax). Independence requires that for all prospects q, r, s, *if* $q \succcurlyeq r$ *then* $(q, p; s, 1 - p) \succcurlyeq (r, p; s, 1 - p)$, for all p. If all three axioms hold, preferences can be represented by

$$V(q) = \bullet\, p_i u(x_i) \tag{1}$$

where q is any prospect, and $u(\cdot)$ is a "utility" function defined on the set of consequences.

The concept of risk is pervasive in economics, so economists naturally need a theory of individual decision making under risk. EU has much to recommend itself in this capacity. The theory has a degree of intuitive appeal. It seems almost trivially obvious that any satisfactory theory of decision making under risk will necessarily take account of both the consequences of choices and their associated probabilities. These are, by definition, the dimensions relevant in the domain of risk. EU provides one very simple way of combining probabilities and consequences into a single "measure of value," which has a number of appealing properties. One such property is monotonicity, which can be defined as follows: Let $x_1, \ldots, x_n$ be consequences ordered from worst (x_1) to best (x_n). We may say that one prospect $q = (p_{q1}, \ldots, p_{qn})$ first-order stochastically dominates another prospect $r = (p_{r1}, \ldots, p_{rn})$ if for all $i = 1, \ldots, n$,

$$\underset{j=i}{\overset{n}{\bullet}}\ p_{qj} \geq \underset{j=i}{\overset{n}{\bullet}}\ p_{rj} \tag{2}$$

with a strict inequality for at least one *i*. Monotonicity is the property where by stochastically dominating prospects are preferred to prospects that they dominate and it is widely held that any satisfactory theory—descriptive or normative—should embody monotonicity. I will have more to say about this later.

The shape of the utility function also has a simple behavioral interpretation whereby concavity (convexity) of $u(\cdot)$ implies risk averse (prone) behavior; an agent with a concave utility function will always prefer a certain amount x to any risky prospect with expected value equal to x. Modeling risk preferences in this way does collapse some potentially distinct concepts into a single function: any attitude to chance (e.g., like or dislike of taking risks) and any attitude toward consequences (e.g., a diminishing marginal utility of money) must all be captured by the utility function. That need not imply any weakness of the theory. Indeed it is precisely the simplicity and economy of EU that has made it such a powerful and tractable modeling tool. My concern, however, is with the descriptive merits of the theory and, from this point of view, a crucial question is whether EU provides a sufficiently accurate representation of actual choice behavior. The evidence from a large number of empirical tests has raised some real doubts on this score.

3. Descriptive Limitations of Expected Utility Theory—The Early Evidence

Empirical studies dating from the early 1950s have revealed a variety of patterns in choice behavior that appear inconsistent with EU. I shall not attempt a full-blown review of this evidence.[3] Instead, I discuss one or two examples to illustrate the general nature of this evidence, and offer a discussion of its role in stimulating the development of new theories. With hindsight, it seems that violations of EU fall under two broad headings: those that have possible explanations in terms of some "conventional" theory of preferences and those that apparently do not. The former category consists primarily of a series of observed violations of the independence axiom of EU; the latter, of evidence that seems to challenge the assumption that choices derive from well-defined preferences. Let us begin with the former.

There is now a large body of evidence that indicates that actual choice behavior may *systematically* violate the independence axiom. Two examples of such phenomena, first discovered by Maurice Allais (1953), have played a particularly important role in stimulating and shaping theoretical developments in non-EU theory. These are the so-called *common consequence effects* and *common ratio effects*. The first sighting of such effects came in the form of the following pair of hypothetical choice problems. In the first you have to imagine choosing between the two prospects: s_1 = ($1M,1) or r_1 = ($5M, 0.1; $1M, 0.89; 0, 0.01). The first

[3] Those interested in more thorough reviews are recommended to consult Schoemaker (1982) and, more recently, Camerer (1995).

option gives one million U.S. dollars for sure; the second gives five million with a probability of 0.1; one million with a probability of 0.89, otherwise nothing.[4] What would you choose? Now consider a second problem where you have to choose between the two prospects: $s_2 = (\$1M, 0.11; 0, 0.89)$ or $r_2 = (\$5M, 0.1; 0, 0.9)$. What would you do if you really faced this choice?

Allais believed that EU was not an adequate characterization of individual risk preferences and he designed these problems as a counterexample. As we shall shortly see, a person with expected utility preferences would either choose both "*s*" options, or choose both "*r*" options across this pair of problems. He expected that people faced with these choices might opt for s_1 in the first problem, lured by the certainty of becoming a millionaire, and select r_2 in the second choice, where the odds of winning seem very similar, but the prizes very different. Evidence quickly emerged that many people did respond to these problems as Allais had predicted. This is the famous "Allais paradox" and it is one example of the more general common consequence effect.

Most examples of the common consequence effect have involved choices between pairs of prospects of the following form: $s^* = (y, p; c, 1 - p)$ and $r^* = (q, p; c, 1 - p)$, where $q = (x, \bullet; 0, 1 - \bullet)$ and $0 < \bullet < 1$.[5] The payoffs c, x, and y are nonnegative (usually monetary) consequences such that x x y. Notice that both prospects s^* and r^* give outcome c with probability $1 - p$: this is the "common consequence" and it is an obvious implication of the independence axiom of EU that choices between s^* and r^* should be independent of the value of c.[6] Numerous studies, however, have found that choices between prospects with this basic structure are systematically influenced by the value of c. More specifically, a variety of experimental studies[7] reveal a tendency for individuals to choose s^* when $c = y$, and r^* when $c = 0$.

A closely related phenomenon, also discovered by Allais, is the so called *common ratio effect*. Suppose you had to make a choice between \$3000 for sure, or entering a gamble with an 80% chance of getting \$4000 (otherwise nothing). What would you choose? Now think about what you would do if you had to choose either a 25% chance of gaining \$3000 or a 20% chance of gaining \$4000. A good deal of evidence suggests that many people would opt for the certainty of \$3000 in the first choice and opt for the 20% chance of \$4000 in the second. Such a pattern of choice, however, is inconsistent with EU and would constitute one example of the common ratio effect. More generally, this phenomenon is observed in choices among pairs of problems with the following form: $s^{**} = (y, p; 0, 1 - p)$ and $r^{**} = (x, \bullet p; 0, 1 - \bullet p)$ where $x > y$. Notice that the ratio of "winning" probabilities ($\bullet$) is constant, and for pairs of prospects of this structure, EU

[4] In Allais's original examples, consequences were French Francs.

[5] It will be convenient to use a scaling factor λ at several points in the paper, so to avoid repetition, assume $0 < \lambda < 1$ throughout.

[6] The original Allais problems are recovered from this generalization setting $x = \$5M$; $y = \$1M$, $p = 0.11$ and $\lambda = 10/11$.

[7] Examples include H. Moskowitz (1974), Paul Slovic and Amos Tversky (1974), and MacCrimmon and Larsson (1979).

implies that preferences should not depend on the value of p.[8] Yet numerous studies reveal a tendency for individuals to switch their choice from s^{**} to r^{**} as p falls.[9]

It would, of course, be unrealistic to expect any theory of human behavior to predict accurately one hundred percent of the time. Perhaps the most one could reasonably expect is that departures from such a theory be equally probable in each direction. These phenomena, however, involve systematic (i.e., predictable) directions in majority choice. As evidence against the independence axiom accumulated, it seemed natural to wonder whether assorted violations of it might be revealing some underlying feature of preferences that, if properly understood, could form the basis of a unified explanation. Consequently, a wave of theories designed to explain the evidence began to emerge at the end of the 1970s. Most of these theories have the following features in common: (i) preferences are represented by some function $V(\cdot)$ defined over individual prospects; (ii) the function satisfies ordering and continuity; and (iii) while $V(\cdot)$ is designed to permit observed violations of the independence axiom, the principle of monotonicity is retained. I will call theories with these properties *conventional theories*. The general spirit of the approach is to seek "well-behaved" theories of preference consistent with observed violations of independence: I call this general approach the *conventional strategy*.

There is evidence to suggest that failures of EU may run deeper than violations of independence. Two assumptions implicit in any conventional theory are *procedure invariance* (preferences over prospects are independent of the method used to elicit them) and *description invariance* (preferences over prospects are purely a function of the probability distributions of consequences implied by prospects and *do not depend on how those given distributions are described*). While these assumptions probably seem natural to most economists—so natural that they are rarely even discussed when stating formal theories—there is ample evidence that, in practice, both assumptions fail.

One well-known phenomenon, often interpreted as a failure of procedure invariance, is *preference reversal*. The classic preference reversal experiment requires individuals to carry out two distinct tasks (usually separated by some other intervening tasks). The first task requires the subject to choose between two prospects: one prospect (often called the \$-bet) offers a small chance of winning a "good" prize; the other (the "P-bet") offers a larger chance of winning a smaller prize. The second task requires the subject to assign monetary values—usually minimum selling prices denoted M(\$) and M(P)—to the two prospects. Repeated studies have revealed a tendency for individuals to chose the P-bet (i.e., reveal

[8] To see why, consider any pair of options (s_1^{**}, r_1^{**}) where $p = p_1$, then define a further pair of options (s_2^{**}, r_2^{**}) identical except having a *lower* value of $p = p_2$. Since there must be some λ, $(1 > \lambda > 0)$, such that $p_2 = \lambda\, p_1$, we can write $s_2^{**} = (s_1^{**}, \lambda; 0, 1 - \lambda)$ and $r_2^{**} = (r_1^{**}, \lambda; 0, 1 - \lambda)$. It then follows directly from independence that choices between such pairs of prospects should not depend on the value of p.

[9] Examples include Loomes and Sugden (1987), Starmer and Sugden (1989), and Raymond Battalio, Kagel, and Jiranyakul (1990).

P × $) while placing a higher value on the $-bet—i.e., M($) > M(P).[10] This is the so called *preference reversal phenomenon* first observed by psychologists Sarah Lichtenstein and Paul Slovic (1971) and Harold Lindman (1971). It presents a puzzle for economics because, viewed from the standard theoretical perspective, both tasks constitute ways of asking essentially the same question, that is, “which of these two prospects do you prefer?” In these experiments, however, the ordering revealed appears to depend upon the elicitation procedure.

One explanation for preference reversal suggests that choice and valuation tasks may invoke different mental processes that in turn generate different orderings of a given pair of prospects (see Slovic 1995). Consequently, the rankings observed in choice and valuation tasks cannot be explained with reference to a *single* preference ordering. An alternative interpretation explains preference reversal as a failure of transitivity (see Loomes and Sugden 1983): assuming that the valuation task reveals true monetary valuations, (i.e., M($) ~ $; M(P) ~ P), preference reversal implies P > $ ~ M($) > M(P) ~ P; which involves a violation of transitivity (assuming that more money is preferred to less). Although attempts have been made to explain the evidence in ways that preserve conventional assumptions—see, for example, Holt (1986); Karni and Safra (1987); Segal (1988)—the weight of evidence suggests that failures of transitivity and procedure invariance both contribute to the phenomenon (Loomes, Moffat, and Sugden 1998; Tversky, Slovic, and Kahneman 1990).

There is also widespread evidence that very minor changes in the presentation or “framing” of prospects can have dramatic impacts upon the choices of decision makers: such effects are failures of description invariance. Here is one famous example by Tversky and Kahneman (1981) in which two groups of subjects—call them groups I and II—were presented with the following cover story:

> Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

Each group then faced a choice between two policy options:

OPTIONS PRESENTED TO GROUP I:

If program A is adopted, 200 people will be saved

If program B is adopted, there is a 1/3 probability that 600 people will be saved, and a 2/3 probability that no people will be saved.

OPTIONS PRESENTED TO GROUP II:

If program C is adopted, 400 people will die.

If program D is adopted, there is a 1/3 probability that nobody will die, and a 2/3 probability that 600 people will die.

[10] Reviews of this evidence are contained in Tversky and Thaler (1990), Hausman (1992), and Tammi (1997).

The two pairs of options are stochastically equivalent. The only difference is that the group I description presents the information in terms of *lives saved* while the information presented to group II is in terms of lives lost. Tversky and Kahneman found a very striking difference in responses to these two presentations: 72% of subjects preferred option A to option B while only 22% of subjects preferred C to D. Similar patterns of response were found among groups of undergraduate students, university faculty, and practicing physicians.

Failures of procedure invariance and description invariance appear, on the face of it, to challenge the very idea that choices can, in general, be represented by *any* well-behaved preference function. If that is right, they lie outside the explanatory scope of the conventional strategy. Some might even be tempted to say that choices lie outside the scope of economic theory altogether. That stronger claim, however, is controversial, and I will not be content to put away such challenging evidence so swiftly. For present purposes, let it suffice to make two observations. First, whether or not we have adequate economic theories of such phenomenon, the "Asian disease" example is clearly suggestive that framing effects have a bearing on issues of genuine economic relevance. Second, there are at least some theories of choice that predict phenomena like preference reversal and framing effects, and some of these models have been widely discussed in the economics literature. Although most of these theories—or at least the ones I will discuss—draw on ideas about preference to explain choices, they do so in unorthodox ways, and many draw on concepts more familiar to psychologists than economists. The one feature common to this otherwise heterodox bunch of theories is that none of them can be reduced to or expressed purely in terms of a single preference function $V(\cdot)$ defined over individual prospects. I will call such models *nonconventional theories*. These theories step into what has been relatively uncharted water for the economics profession. One of the aims of this chapter will be to reflect on the relative merits of the conventional and nonconventional approaches.

4. Nonexpected Utility Theories

4.1. The Conventional Strategy

One way to approach this literature is to ask a question that motivated a number of theories: what properties would a conventional theory of preference need to explain the known violations of independence? To pursue that question, it will be helpful to introduce an expositional device known as the probability triangle diagram,[11] this will also prove useful as a vehicle for comparing the predictions of alternative theories.

[11] Although the probability triangle had appeared in the literature many years before (see Marschak 1950), Mark Machina's use of it in the 1980s (see further on) popularized it to the extent that some have called this diagram the "Machina triangle."

Consider the class of prospects defined over three outcomes x_1; x_2; x_3 such that $x_1 \succ x_2 \succ x_3$. Since any such prospects can be described as a vector of probabilities (p_1, $1 - p_1 - p_3$, p_3) we can also locate them, graphically, in two-dimensional probability space. Figure 4.1a is a probability triangle that does this for the four prospects $\{s_1, r_1, s_2, r_2\}$ from the original Allais paradox problems. By convention, the horizontal axis measures the probability of the worst consequence ($0) increasing from left to right; the vertical axis measures the probability of the best consequence ($5M) increasing from bottom to top. Hence s_1, which results in the intermediate consequence of $1M for sure, is located at the bottom left corner of the triangle; s_2 and r_2, which each assign positive probability to only two of the three possible consequences, are located on the triangle boundaries; while r_1, which assigns positive probability to all three consequences, lies on the interior of the triangle. Two lines have been drawn in the triangle joining the pairs of prospects involved in the two choices. It is easy to establish that these two lines are parallel.

Taking into consideration ordering plus continuity, we can see that preferences over prospects in any given triangle can be represented by a set of indifference curves. Hence, every conventional theory implies the existence of a set of indifference curves in this space though the precise form of indifference curves varies between them.

The addition of the independence axiom of EU restricts the set of indifference curves to being *upward sloping* (left to right), *linear*, and *parallel*. One such set of indifference curves is illustrated in figure 4.1b (preferences are increasing moving north-west). Independence is a strong restriction that leaves only one feature of

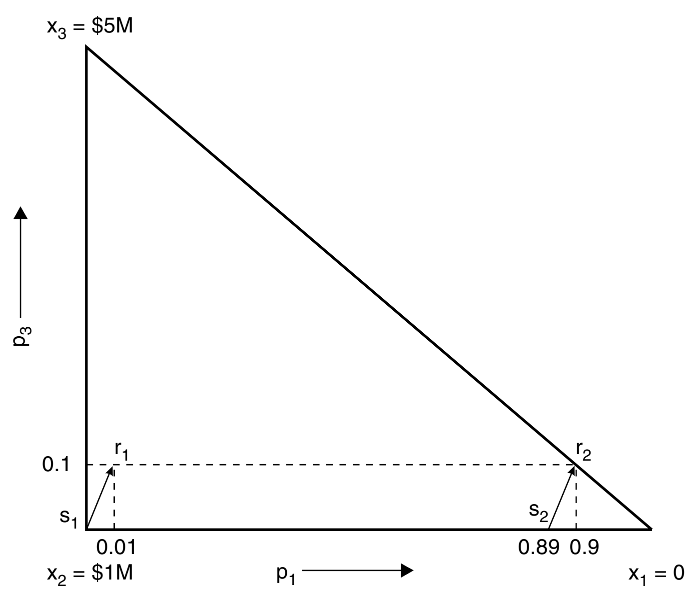


FIGURE 4.1a The Allais paradox problems in a probability triangle.

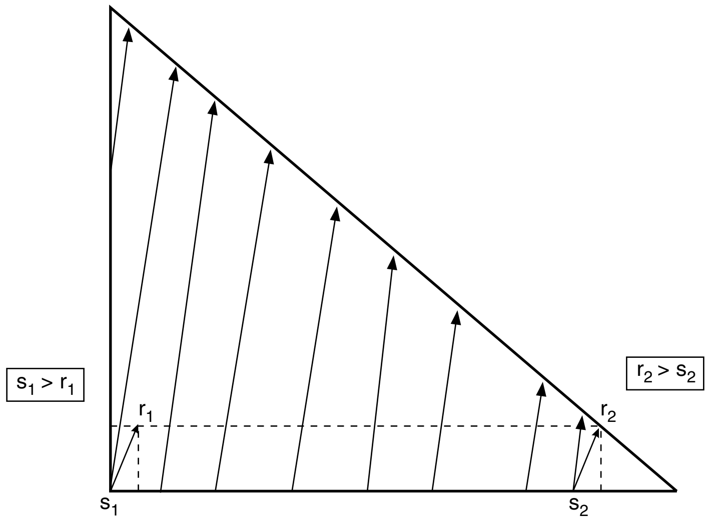


FIGURE 4.1b Expected-utility indifference curves.

the indifference curves undetermined; that is, their slope. In EU, the slope of the indifference curves reflects attitude to risk and may vary among individuals: the more risk averse the individual, the steeper the slope of his or her indifference curves. To see why, look at figure 4.1c and consider two individuals: person 1 has indifference curves with the slope of the dashed line (hence $s \sim r$); person 2 has indifference curves with the slope of the solid line (hence $s \sim r'$). Person 2 can be seen to be the more risk averse in the sense that, as we move northwest along the hypotenuse, relative to person one, we must give her a higher chance of winning the best outcome in the riskier prospect in order to generate indifference with the safe prospect s.

In relation to the Allais paradox problems in figure 4.1b, for a given individual, EU allows three possibilities. Indifference curves could have a steeper slope than the lines connecting prospects, in which case $s_1 \succ r_1$ and $s_2 \succ r_2$. This is the case represented in figure 4.1b. Alternatively, indifference curves could have a less steep slope (in which case $r_1 \succ s_1$ and $r_2 \succ s_2$). Finally, the slope of indifference curves could correspond exactly with that of the lines joining pairs of prospects, in which case $r_1 \sim s_1$ and $r_2 \sim s_2$. But as noted above, people often violate EU, revealing $s_1 \succ r_1$ in the left-hand problem, $r_2 \succ s_2$ in the right-hand problem. Relative to the predictions of EU, in choosing r_2 over s_2 these people are being more risk-seeking than they should be, given their choice of s_1 over r_1.

A similar tendency is apparent in the common ratio effect. A pair of common ratio problems is illustrated in figure 4.2. The pair of prospects $\{s_1^{**}, r_1^{**}\}$, near the left edge of the triangle, corresponds with the common ratio problems where $p = 1$. As p falls, we generate pairs of prospects like $\{s_2^{**}, r_2^{**}\}$ located on parallel lines further to the right in the triangle. Assuming expected utility preferences,

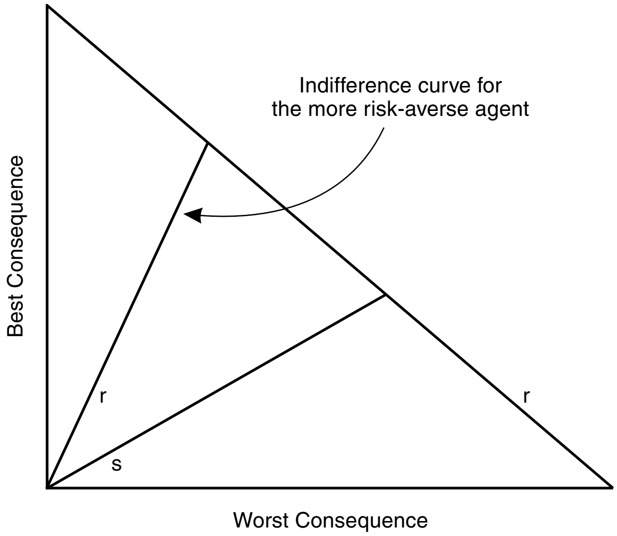


FIGURE 4.1c Different degrees of risk-aversion in EU.

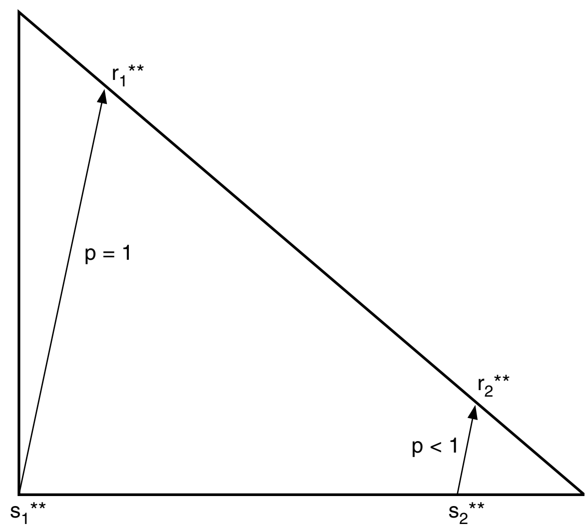


FIGURE 4.2 Common ratio prospects.

an individual must either prefer the "safer option" in *both* choices *or* the "riskier option" in *both* choices, yet many people choose s_1^{**} over r_1^{**} and s_2^{**} over r_2^{**}. This is the common ratio effect and, as in the common consequence effect, relative to the predictions of EU, there is an "inconsistency" in the risk attitudes revealed across their choices.

Viewed in the context of the triangle, this inconsistency is suggestive of a systematic pattern: relative to the predictions of EU, choices between prospects located in the bottom right-hand corner appear more risk-prone than should be expected given preferences revealed for choices located leftward and/or upward in the triangle. Any conventional theory seeking to explain these standard violations of EU will therefore need at least one quite specific property: indifference curves determining preferences over pairs of prospects located near the right-hand corner of a given triangle—e.g., $\{s_2^{**}, r_2^{**}\}$—will need to be relatively flat (reflecting more risk-prone behavior), compared with indifference curves determining choices over pairs of prospects, like $\{s_1^{**}, r_1^{**}\}$, near to the left-hand edge of the triangle. All of the proposed conventional alternatives to EU are able to generate this property, though they do so in a variety of ways.

4.1.1. THE "FANNING-OUT" HYPOTHESIS

Having observed this apparent connection among different violations of independence, Mark Machina (1982) proposed an analytical extension of EU (termed "generalized expected utility analysis"), along with a specific hypothesis on the shape of nonexpected utility indifference curves. Analytically, he noted that under expected utility, where $V(q) = \text{í}U(x_i)p_i$, the utility values $U(x_i) = \bullet V(q)/\bullet p_i$ are the probability derivatives of $V(\cdot)$. He then showed that standard expected utility results (e.g., risk aversion = concavity of $U(\cdot)$) also hold for the probability derivatives $U(x_i; q) = \bullet V(q)/\bullet p_i$ of smooth nonexpected utility preference functions $V(\cdot)$, so that $U(\cdot;q)$ can be thought of as the "local utility function" of $V(\cdot)$ about q. For example, the property "concavity of $U(\cdot;q)$ at every q" is equivalent to global risk aversion of $V(\cdot)$.

Given the existence of phenomena like the common ratio and common consequence effects, Machina hypothesized that the local utility functions $U(\cdot; q)$ become more concave as we move from (first order) stochastically dominated to stochastically dominating distributions. Loosely speaking, this essentially empirical assumption (which Machina calls "Hypothesis II") implies a tendency for agents to become more risk averse as the prospects they face get better; in the context of the triangle, it means that indifference curves become steeper, or "fan out," as we move northwest. Figure 4.3 illustrates the general pattern of indifference curves implied by Hypothesis II. Notice that they are drawn as wavy lines: generalized expected utility theory requires indifference curves to be smooth but does not imply that they must be linear (though they may be). It is very easy to see that this fanning-out property generates implications consistent with the common consequence and common ratio effects. Since indifference curves are relatively steeply sloped in the neighborhood of prospect m, m lies on a higher indifference curve than q or r. Flatter indifference curves in the bottom right-hand corner of the triangle are such that t lies on a higher indifference curve than s. Hence, for an individual whose indifference curves fan out we can construct prospects over which we will observe a common consequence effect (e.g., $m \succ q$ and $t \succ s$) and a common ratio effect (e.g., $m \succ r$ and $t \succ s$).

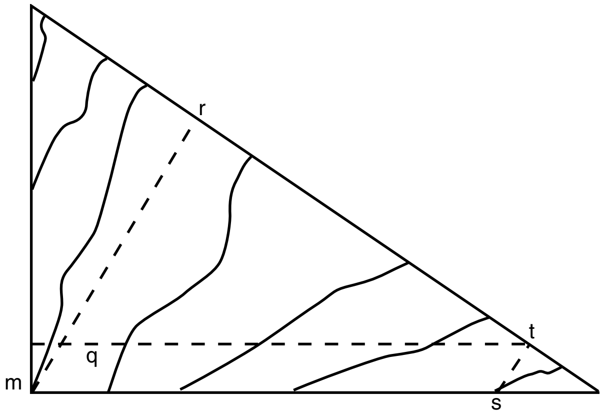


FIGURE 4.3 Indifference curves in generalized expected utility.

A whole family of models have this fanning-out property and, within this family, one important subset consists of those models that restrict indifference curves to be linear. One example is Soo Hong Chew and Kenneth MacCrimmon's (1979) weighted-utility theory in which preferences over prospects are represented by the function:

$$V(q) = [\bullet\, p_i \cdot g(x_i) \cdot u(x_i)]/[\bullet\, p_i \cdot g(x_i)], \tag{3}$$

where $u(\cdot)$ and $g(\cdot)$ are two different functions assigning non-zero weights to all consequences. The model incorporates EU as the special case in which the weights assigned by $g(\cdot)$ are identical for every consequence. Weighted utility has been axiomatized by, among others, Chew and MacCrimmon (1979a), Chew (1983), and Fishburn (1983), and different variants are discussed in Fishburn (1988). Essentially these axiomatizations involve a weakened form of the independence axiom, which constrains indifference curves to be linear without requiring them to be parallel. One version of weak independence is this: *if* $q \succ r$ then for each p_q there exists a corresponding p_r such that $(q, p_q; s, 1 - p_q) \succ (r, p_r; s, 1 - p_r)$ for all s. If we think in terms of preferences in the triangle diagram, excepting the special case of EU, this axiom has the effect of requiring there to be some point at which all indifference curves cross. The location of this point, which could lie inside or outside of the triangle boundary, depends upon the specifications of the functions $u(\cdot)$, and $g(\cdot)$. Transitivity can be preserved by making the point from which curves radiate lie outside the boundary of the triangle and, to explain the common ratio and common consequence effects, the origin of indifference curves must lie somewhere to the southwest of the triangle, as in figure 4.4. Having restricted the model in this way,[12] we can then understand it as a special

[12] Chew and MacCrimmon (1979b) explain the conditions necessary to generate this property.

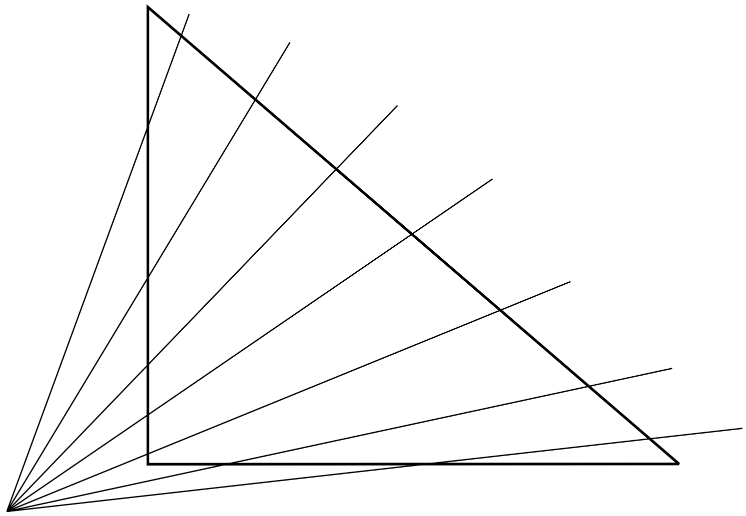


FIGURE 4.4 Weighted-utility theory with indifference curves panning out.

case of Machina's theory (including Hypothesis II), in which indifference curves are constrained to be linear.

It is not obvious to me that weak independence has much, if any, intuitive appeal, and the main rationale for assuming it in weighted utility theory is presumably that it results in a simple mathematical function capable of generating fanning out and hence explaining the early violations of EU. Other models with very similar properties have been based on psychologically grounded hypotheses. One example is the theory of disappointment developed by Bell (1985) and Loomes and Sugden (1986). While this theory lacks axiomatic foundations, it has a more obvious intuitive interpretation. In the version presented by Loomes and Sugden, preferences over prospects can be represented by the function

$$V(q) = \bullet_i \, p_i \, [u(x_i) + D(u(x_i) - \bullet)], \tag{4}$$

where $u(x_i)$ is interpreted as a measure of "basic" utility (that is, the utility of x_i, considered in isolation from the other consequences of q) and $\bullet$ is a measure of the "prior expectation" of the utility from the prospect. The model assumes that if the outcome of prospect is worse than expected (i.e., if $u(x_i) < \bullet$) a sense of disappointment will be generated. On the other hand, an outcome better than expected will stimulate "elation." With $D(\cdot) = 0$, the model reduces to EU. This additional function, however, is intended to capture a particular intuition about human psychology: that people dislike disappointment and so act to avoid it. More specifically, this is captured by assuming that agents are "disappointment averse" ($D(h)$ is concave for $h < 0$) and "elation prone" ($D(\cdot)$ is convex for $h > 0$). The theory then implies a tendency for indifference curves to fan out in the triangle. The theory of disappointment has close affinity with earlier models based on *moments of*

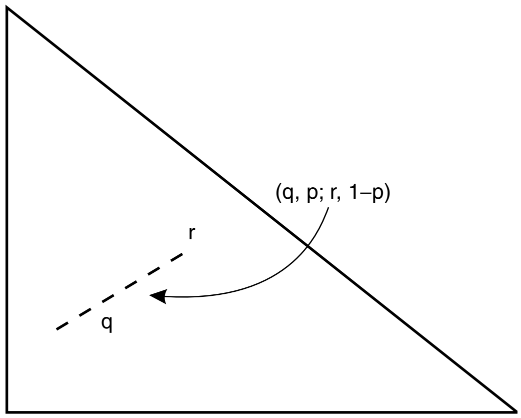


FIGURE 4.5a Probability mixtures of prospects q, r.

utility. In EU, the value of a prospect is the (probability-weighted) mean of utility. Allais (1979) proposed a model in which $V(\cdot)$ may also depend on the second moment of utility, that is, the variance of utility about the mean. Hagen (1979) extended this idea to include the third moment of utility, or skewness. Sugden (1986) shows that properties of $D(\cdot)$ imposed in disappointment theory can be interpreted as restrictions on Hagen's general model of moments.

A series of other models with linear indifference curves including implicit expected utility (Dekel 1986) and implicit weighted utility (Chew 1989) allow fanning out, but also permit more complex patterns. For example, Faruk Gul (1991) and William Neilson (1992) present models based on implicit expected utility that generate a mixture of fanning-in and fanning-out within a given triangle.[13] The crucial axiom in these models is a weakened form of independence called "betweenness": if $q \succ r$, then $q \succ (q, p; r, (1 - p)) \succ r$ for all $p < 1$. It is this assumption that imposes linearity on indifference curves, and, conversely, it is implied by any model that assumes linear indifference curves.

Behaviorally, betweenness implies that any probability mixture of two lotteries will be ranked between them in terms of preference, and, given continuity, an individual will be indifferent to randomization among equally valued prospects. To understand the connection between these behavioral and geometric properties, look at figure 4.5a and consider an individual who is offered a compound gamble giving a p chance of prospect q and a $1 - p$ chance of r. Geometrically, the simple prospect induced by this compound gamble must lie along the straight line joining q and r (for any 0 • p • 1). For an individual with linear indifference curves, it follows that for any $q \sim r$, the indifference curve through q and r coincides with the set of simple prospects induced by $(q, p; r, 1 - p)$. Hence, with linear indifference curves, the individual indifferent between q and r is also indifferent to

[13] These models were proposed in response to later evidence (see section 5) that suggests behavior is more complex than pure fanning-out theories imply.

randomization between them. Once betweenness is relaxed, this indifference to randomization no longer holds, and two important cases can be distinguished: quasi-convex preferences and quasi-concave preferences. A preference function is strictly quasi-convex if for every $q \bullet r$, $V(q, p; r, (1 - p)) < \max[V(q), V(r)]$ for all p. When preferences are quasi-convex, indifference curves are *concave*, as in figure 4.5b, and consequently the individual will be averse to randomization among equally valued prospects (notice that prospects r and s in figure 4.5b lie on a higher indifference curve than probability mixtures of the two prospects that lie along the dashed line). Conversely, when preferences are strictly quasi-concave, indifference curves are *convex*, as in figure 4.5c, hence, by similar reasoning, individuals

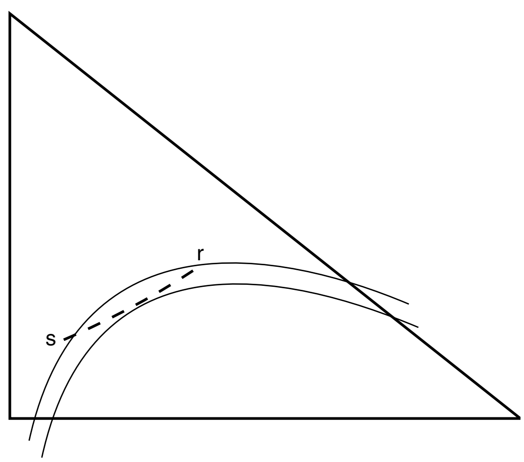


FIGURE 4.5b Quasi-convex preferences. Aversion to randomization.

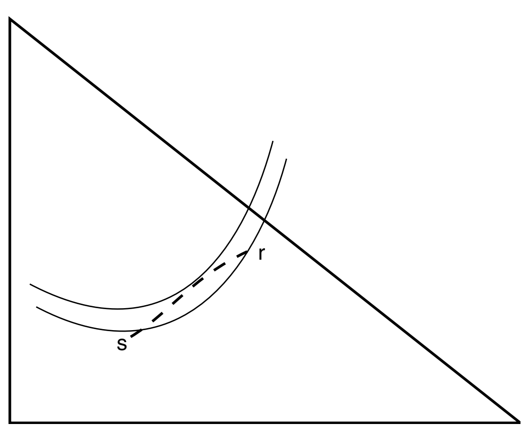


FIGURE 4.5c Quasi-concave preferences. Preference for randomization.

prefer to randomize among equally valued prospects. Some significant theoretical results in economics extend to a nonexpected utility world if agents' preferences satisfy betweenness (see Section 5.3).

Various models have been proposed that do not impose betweenness. Chew, Epstein, and Segal (1991) propose *quadratic utility theory*, which relies on a weakened form of betweenness called *mixture symmetry*: if $q \sim r$ then $(q, p; r, (1 - p)) \sim (q, (1 - p); r, p)$. In this model, indifference curves may switch from concave to convex (or vice versa) as we move across the triangle. Joao Becker and Rakesh Sarin (1987) propose a model with even weaker restrictions. Their *lottery-dependent utility* assumes only ordering, continuity, and monotonicity. The basic model is conventional theory for minimalists as, without further restriction, it has virtually no empirical content. The authors discuss a particular "exponential form," which implies fanning out.

An important subset of the betweenness nonconforming theories has an additional feature absent from the models discussed so far. Upto this point we have considered a variety of conventional theories, each of which generates the property of fanning out. Although they achieve it in different ways, there is one structural similarity between these theories: each operates by assigning subjective weights—or utilities—to consequences; the value assigned to any given prospect is then determined by some function that combines these utilities with *objective* probabilities. Another variant of the conventional strategy involves the use of probability transformation functions that convert objective probabilities into subjective *decision weights*. An important feature of these models is that, excepting special cases, betweenness does not hold.

4.1.2. THEORIES WITH DECISION WEIGHTS

There is evidence for the view that individuals have subjective attitudes to probabilities that are distinct from attitudes to consequences. For instance, according to Nick Pidgeon et al. (1992), when people are asked to make judgments about the likelihood of death occurring from different causes, they tend to underestimate the number of deaths from relatively frequent causes, while overestimating deaths due to relatively infrequent causes. Similarly, apparent biases in the subjective odds revealed in studies of racetrack betting have been explained as bettors being either oversensitive to the chances of winning on long shots (Ali 1977; Thaler and Ziemba 1988), or oversensitive to the chances of losing on favorites (Jullien and Salanié 1997). These effects might be revealing misperception of objective probabilities or a tendency for individuals subjectively to weight objective probabilities. Either way, in principle, such effects could be captured in models incorporating *decision weights*. A number of such theories can be understood as variants of the following functional form where the φ_i terms represent decision weights:

$$V(q) = \bullet_i\, \pi_i \cdot u(x_i). \tag{5}$$

I will call this the *decision-weighted form*. Theories of this type were first discussed by Ward Edwards (1955, 1962). In its most basic form, consequences are treated in the way in which probabilities are handled in the standard theory and enter "raw" with $u(x_i) = x_i$ for all i. Edwards called this *subjective expected value*, and in the version presented by Jagdish Handa (1977) the decision weight attached to each outcome is determined by a *probability weighting function* "(p_{i}), which transforms the individual probabilities of each consequence directly into weights. As in most theories that incorporate probability weights, "($\cdot$) is assumed to be increasing with "(1) = 1 and "(0) = 0, and I will retain these assumptions from now on. The subjective expected value form has not been widely used, but theories that allow nonlinear transformations of both probabilities and consequences have received much more attention. In the simplest variant of this latter type of model, individuals are assumed to maximize the function

$$V(q) = \bullet\,(p_i) \cdot u(x_i). \tag{6}$$

I will call this form *simple decision weighted utility*.[14] Both this and subjective expected value, because they transform the probabilities of individual consequences *directly* into weights (i.e., $\pi_i = \bullet\,(p_i)$), have the property that $V(q)$ will not generally satisfy monotonicity. To see this, suppose for the sake of example that $\bullet\,(\cdot)$ is convex, then $\bullet\,(p) + \bullet\,(1 - p) < 1$ and there will be some $\bullet > 0$ such that gambles of the form $(x, p; x + \bullet, 1 - p)$ will be rejected in favor of $(x, 1)$, even though they stochastically dominate the sure option. A similar argument applies for *any* departure from linearity, and the only way to ensure general monotonicity in this type of theory is to set decision weights equal to objective probabilities (i.e., $\pi_i = \bullet\,(p_i) = p_i$ for all i), in which case the theory reduces to EU. This property was first noted by Fishburn (1978) and since then has been widely viewed as a fatal objection to models that attach decision weights to the raw probabilities of individual consequences. For example, Machina (1983, p. 97) argues that any such theory will be, "in the author's view at least, unacceptable as a descriptive or analytical model of behavior." The point seems to have been generally accepted, and, while many theorists have wished to retain the idea that probabilities may be subjectively weighted, the thrust of work in this stream of the literature over the past two decades has been toward variants of the decision-weighting form that satisfy monotonicity.

There are two distinct strands to this contemporary literature: one conventional, the other distinctly nonconventional. The nonconventional route is that taken by Kahneman and Tversky (1979) in *prospect theory*, but that model takes us outside the bounds of conventional theory, and so I postpone further discussion of it until the next section. Theorists following the conventional route have proposed decision-weighting models with more sophisticated probability transformations designed to ensure monotonicity of $V(\cdot)$. One of the best-known models of this type is *rank-dependent expected-utility theory*, which was first proposed by John Quiggin (1982). Machina (1994) describes the rank-dependent model as

[14] This form has sometimes been called *subjective expected utility*, but this label is now more commonly used to refer to L. Savage's (1954) formulation of EU.

"the most natural and useful modification of the classical expected utility formula" and, as testament to this, it has certainly proved to be one of the most popular among economists. In this type of model the weight attached to any consequence of a prospect depends not only on the true probability of that consequence but also on its ranking relative to the other outcomes of the prospect. With consequences indexed as before such that x_1 is worst and x_n best, we can state rank-dependent expected-utility theory as the hypothesis that agents maximize the decision-weighted form with weights given by

$$\pi_i = \bullet\,(p_i + \cdots + p_n) - \bullet\,(p_{i+1} + \cdots + p_n) \quad \text{for } i = 1, \ldots, n-1,$$
$$\pi_i = \bullet\,(p_i) \quad \text{for } i = n.$$

In this model there is a meaningful distinction between decision weights (φ) and probability weights (•). Richard Gonzalez and George Wu (1999, p. 135) suggest an interpretation of the probability-weighting function as reflecting the underlying "psychophysics of risk," that is, the way in which individuals subjectively "distort" objective probabilities; the decision weight then determines how the probability weights enter the value function $V(\cdot)$. Notice that $\bullet\,(p_i + \cdots + p_n)$ is a subjective weight attached to the probability of getting a consequence of x_i or better, and $\bullet\,(p_{i+1} + \cdots + p_n)$ is a weight attached to the probability of getting a consequence better than x_i, hence in this theory $\bullet\,(\cdot)$ is a transformation on cumulative probabilities. This procedure for assigning weights ensures that $V(\cdot)$ is monotonic. It also has the appealing property that, in contrast to the simple decision-weighting models that assign the same decision weight to any consequence with probability p, the weight attached to a consequence may vary according to how "good" or "bad" it is. So in principle this would allow for, say, extreme outcomes to receive particularly high (or low) weights. A less appealing feature of the model is that a small change in the value of some outcome of a prospect can have a dramatic effect on its decision weight if the change affects the rank order of the consequence; but a change in the value of an outcome, no matter how large the change, can have no effect on the decision weight if it does not alter its rank.

The predictions of the rank-dependent model rely crucially on the form of $\bullet\,(\cdot)$. If $\bullet\,(\cdot)$ is convex, this generates a set of concave indifference curves (implying aversion to randomization) that are parallel at the hypotenuse but fan out as we move left to right across the triangle and fan in (i.e., become less steep) as we move vertically upwards. Aside from the hypotenuse parallelism that holds for any $\bullet\,(\cdot)$ (see Camerer 1989), the reverse pattern of indifference curves (i.e., convex curves, horizontal fanning in, and vertical fanning out) is generated with a concave $\bullet\,(\cdot)$.

Curvature of $\bullet\,(\cdot)$ in the rank-dependent model has been interpreted as reflecting "optimism" and/or "pessimism" with respect to probabilities (see Quiggin 1982; Yaari 1987; Diecidue and Wakker 1999). Consider, for example, the prospect $q = (x_1, 0.5; x_2, 0.5)$. Assigning weights to the consequences of q according to the rank-dependent method above gives $\pi_1 = 1 - \bullet\,(0.5)$ and $\pi_2 = \bullet\,(0.5)$. With $\bullet\,(\cdot)$ convex, $\bullet\,(0.5) < 0.5$, hence the weight attached to the lower ranking consequence, x_1, will be higher than the weight attached to the larger consequence. This overweighting of the lower-ranked consequences relative to higher-ranked consequences can be interpreted as a form of pessimism. Pessimism also has a close connection to risk-aversion: a pessimistic agent with a concave $u(\cdot)$ will be universally risk averse; and an agent with a convex utility function can be risk averse if he or she is sufficiently pessimistic (see Chew, Karni, and Safra 1987; Chateauneuf and Cohen 1994).

Although rank-dependent theory does not imply generalized fanning out, the early evidence of EU violation can be explained either by assuming a simple convex • $(\cdot)$ or by more complex specifications. One possibility is the function displayed in figure 4.6, which has “$(p) = p$ for a unique value of $p = p^*$; it is concave below p^* and convex above it, hence "low" probabilities (below p^*) are overweighted. Quiggin (1982) proposes this form with $p^* = 0.5$. He is drawn to this partly because it explains the early violations of EU and partly because it has the appealing property that 50–50 bets will be undistorted by probability weighting. While there is little empirical support for the crossover at $p = 0.5$, research over a period of fifty years, from Malcolm Preston and Phillip Baratta (1948) to Drazen Prelec (1998), lends support to the hypothesis of an (inverted) s-shaped decision-weighting

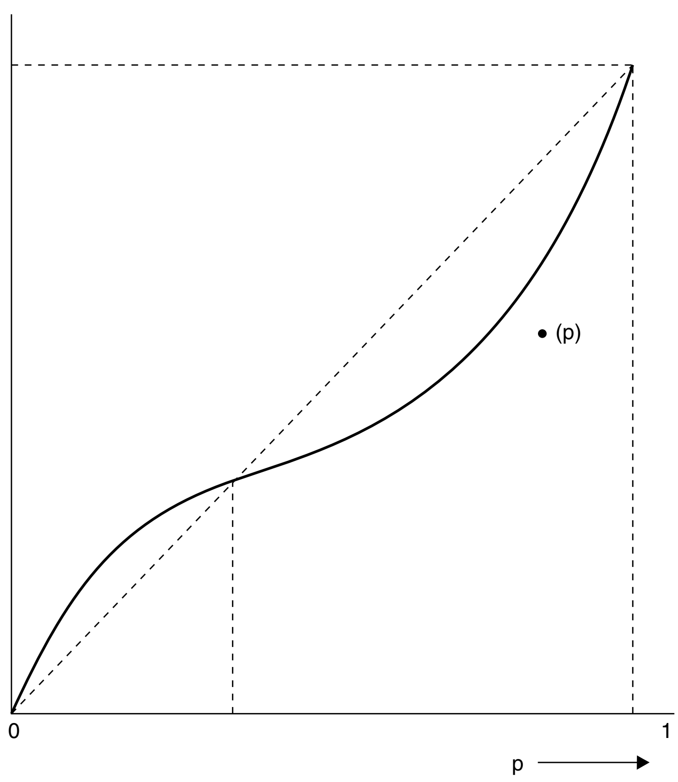


FIGURE 4.6 An (inverted) S-shaped probability weighting function.

function (see section 5.1.1). A useful discussion of the theoretical properties necessary and sufficient for an s-shaped weighting function can be found in Tversky and Wakker (1995).

Axiomatizations of rank-dependent expected utility have been presented by, among others, Segal (1990), Wakker (1994), Abdellaoui (1999), and Yaari (1987), who examine the special case of the model with linear utility (this is essentially a rank-dependent reformulation of Handa's proposal with $u(x_i) = x_i$. Wakker, Erev, and Weber (1994) provide a useful discussion of the axiomatic foundations of rank-dependent expected utility in which they demonstrate the essential difference between EU and rank-dependent expected utility is that the latter theory relies on a weakened form of independence called "comonotonic independence." It is an implication of the standard independence axiom that if two prospects q and r have a common outcome x, which occurs with probability p, in each prospect, substituting x for some other outcome y in both prospects will not affect the preference order of q and r. The same may not be true in the rank-dependent model, however, because such substitutions may affect the rankings of consequences and hence the decision weights. Comonotonic independence asserts that preferences between prospects will be unaffected by substitution of common consequences so long as these substitutions have no effect on the rank order of the outcomes in either prospect.

Various generalizations of the rank-dependent model have been proposed (Segal 1989, 1993; Chew and Epstein 1989; Green and Jullien 1988). In Green and Jullien, the crucial axiom is ordinal independence. Suppose two prospects q, r have a "common tail" such that for some j, $p_{qi} = p_{ri}$ for all i from j to n. Ordinal independence requires that preferences between q and r be unaffected by the substitution of this common tail, in both prospects, with any other common tail. This axiom is necessary for any rank-dependent model. The contribution of Chew and Epstein constructs a theoretical bridge between the rank-dependent models and the betweenness-conforming theories (i.e., those with linear indifference curves discussed previously) by presenting a general model that contains each class as a special case (see also the "correction and comment" by Chew et al. 1993).

A further extension to the rank-dependent model discussed by Starmer and Sugden (1989), Tversky and Kahneman (1992), and Luce and Fishburn (1991) involves a distinction between consequences that are "gains" and those that are "losses." This approach draws on Kahneman and Tversky's earlier work on prospect theory. It is to this model that we now turn, and in doing so we cross the boundary into nonconventional territory.

4.2. Nonconventional Theories

4.2.1. THE PROCEDURAL APPROACH AND REFERENCE DEPENDENCE

Each of the theories we have considered so far models choice as preference maximization and assumes that agents behave *as if* optimizing some underlying preference function. The "as if" is significant here: the conventional approach, interpreted descriptively, seeks to predict which choices *are* made, and typically, there is no presupposition that the model corresponds with any of the mental activities actually involved in making choices. While this underlying methodology dominates economic theory, another approach more common in the psychology literature seeks to model the processes that lead to choice. I will call such theories *procedural theories*. A common feature of such theories is to assume that agents draw on decision heuristics or rules of one kind or another when making their choices. The problem is then to identify the set of decision heuristics that the agent may draw on, and to specify the conditions under which particular rules will be followed. In such theories, it is common for problem context to be an important determinant of choice-rule selection. For instance, there may be a tendency to choose the rule that is easiest to apply in the given context, and ease of application may depend on how a problem is presented. Consequently, it seems natural to expect phenomena like framing effects within this framework.

One recent and quite general procedural model has been developed by John Payne, James Bettman, and Eric Johnson (1993). They assume that agents have at their disposal a range of possible choice-heuristics that might be applied to a given decision task. These include expected utility calculations, satisficing rules, lexicographic choice rules, and so on.[15] In their *adaptive* model the decision maker "decides how to decide," trading off the desire to make a "good" decision against the cognitive effort involved in applying different rules in a given context. Here, as in other procedural models, the agent is conceived of as *boundedly rational*; an agent with limited computational ability and, perhaps, imperfectly defined objectives, attempting to cope with an often complex decision environment. Yet, boundedly rational does not equate with dumb. Payne, Bettman, and Johnson argue that selection of choice procedures is "adaptive and intelligent" (p. 14), and though decisions may not be optimal in the conventional sense, the selection of decision rule does involve optimization but with unusual constraints (e.g., information-processing capacity) and/or objectives (e.g., the choice of strategy might be influenced by considerations such as a desire to be able to justify a choice to a third party). Indeed, as John Conlisk (1996, p. 672) points out, "bounded rationality is not a departure from economic reasoning, but a needed extension of it."

While models of bounded rationality have been applied with some success elsewhere in economics—see Conlisk's (1996) review—full-blown procedural models of decision under risk, like that of Payne, Bettman, and Johnson, have not received much attention from the economics profession. Nevertheless, there has been a degree of cross-fertilization, and some theories involving a procedural element have appeared in the economics literature. Examples include the models proposed by Kahneman and Tversky (1979), Rubinstein (1988), and Lavoie (1992).

The most widely discussed of these is Kahneman and Tversky's (1979) *prospect theory*. In this theory, choice is modeled as a two-phase process. In the first phase,

[15] For a discussion of satisficing rules, see Simon (1955); and for an example of a lexicographic procedure, see Tversky (1969).

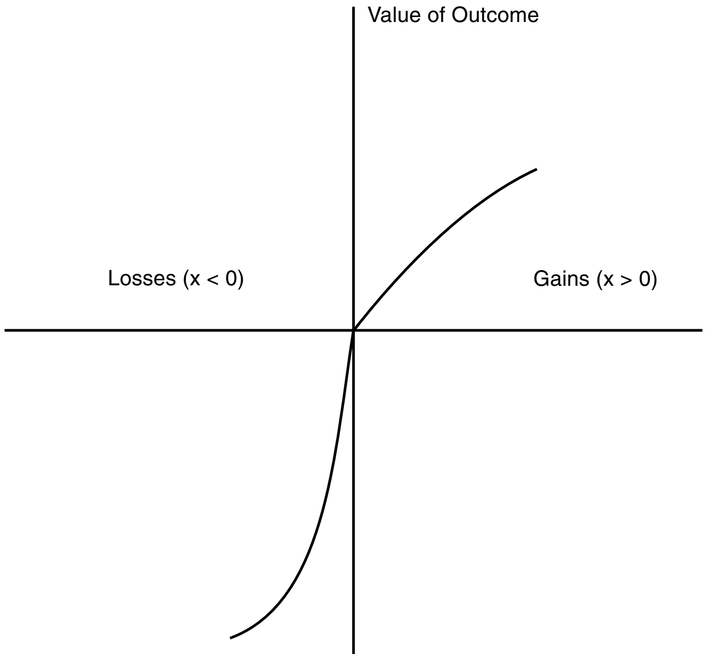


FIGURE 4.7 The valuation of outcomes in prospect theory.

prospects are "edited" using a variety of decision heuristics; in the second, choices among edited prospects are determined by a preference function that, for a restrictive class of prospects,[16] can be represented by the simple decision-weighted utility form defined previously in expression 6. Two features of this theory distinguish it clearly from any of the theories we have discussed so far. First and most obvious is the editing phase, but a second distinguishing feature is that, in prospect theory, outcomes are interpreted as gains and losses relative to a *reference point*. For present purposes we may think of the reference point as status quo wealth. The motivation for handling consequences in this way is that it allows gains and losses to be evaluated quite differently. This capacity, it turns out, has some quite interesting implications.

In prospect theory outcomes are evaluated via a utility function[17] with the shape of that in figure 4.7. It is kinked at the reference point (i.e., status quo, $x = 0$). Notice

[16] Prospect theory does not provide a general preference representation over prospects. Strictly speaking, it applies only to prospects of the form $(x_1, p_1; x_2, p_2; 0, (1 - p_1 - p_2))$. The function assumed in prospect theory coincides with the function defined here in the case of "regular prospects" where either $p_1 + p_2 < 1$, or $x_1 \geq 0 \geq x_2$, or $x_1 \leq 0 \leq x_2$.

[17] Kahneman and Tversky explicitly avoid using the term "utility" to describe this function, preferring instead the term "value function." I suspect they had in mind a conception of value independent of risk and wished to distance themselves from the notion of utility in EU, where utilities may partly reflect attitudes to chance. Here I revert to utility terminology, but with a timely reminder that the appropriate interpretation of "utility" varies among theories.

two further properties: (i) it is concave for gains and convex for losses, and (ii) it is steeper in the domain of losses. In their later paper, Tversky and Kahneman (1992) interpret these restrictions as implications of two more general properties of perception and judgment: *diminishing sensitivity* and *loss-aversion*. Diminishing sensitivity holds that the psychological impact of a marginal change will decrease as we move further away from a reference point. So, for example, relative to the status quo, the difference between a gain of \$10 and \$20 will seem larger than the difference between gains of \$110 and \$120. More generally, the assumption of diminishing sensitivity applied to the outcome domain entails diminishing marginal utility for gains (i.e., $u''(x) \bullet 0$ for $x \bullet 0$) and diminishing marginal disutility for losses (i.e., $u''(x) \bullet 0$ for $x \bullet 0$). So property (*i*) of the utility function is a direct implication of diminishing sensitivity. Loss-aversion is the principle that "losses loom larger than corresponding gains" (Tversky and Kahneman 1992, p. 303). They justify this second feature of the function partly by an appeal to intuition and partly to empirical evidence (e.g., the fact that most people find symmetric bets of the form $(x, 0.5; -x, 0.5)$ "distinctly unattractive"). Loss-aversion is modeled by imposing $u'(x) < u'(-x)$.

The evaluation of risky prospects involves a probability-weighting function and, in the original version of prospect theory, Kahneman and Tversky proposed a weighting function that underweights "large" and overweights "small" probabilities. The endpoints are such that $\bullet(1) = 1$ and $\bullet(0) = 0$, but the function is not defined for probabilities close to zero and one; unusual things may happen in these regions—for example, "very small" probabilities might be ignored. It is worth noting that in a later version of prospect theory (see cumulative prospect theory further on), Kahneman and Tversky adopt the widely used inverted-s weighting function. This is partly because that specification fits their data well, and no doubt partly to resolve the ambiguity about what happens at the end points in the original version, but there is also an underlying theoretical rationale. The principle of diminishing sensitivity, which determines some of the important characteristics of the utility function, can also provide a psychological rationale for an (inverted) s-shaped probability-weighting function: a function with the property of diminishing sensitivity will be steepest close to a reference point, hence on the assumption that the end points of the probability scale constitute natural reference points, diminishing sensitivity implies a probability weighting function that is steep near zero and one but relatively flat around the middle. The inverted-s has precisely these properties. Hence, if diminishing sensitivity is a general principle of perception, it provides a common psychological underpinning for properties of both the utility function and the probability-weighting function.

Kahneman and Tversky (1979) argue that their theory is able to capture a wide range of observed behavior toward risk, including standard violations of the independence axiom (e.g., the common ratio and common consequence effects), and a variety of field data, plus an extensive range of data generated from their own experiments. The theory also has some unusual properties, one of which is the so called *reflection effect*. The fact that concavity of the utility function in the domain of gains is mirrored by convexity in the domain of losses means behavior toward risk can be likewise mirrored across the two domains. For instance, a given individual who displays risk-aversion in a choice among particular prospects with nonnegative outcomes may display risk-seeking if all outcomes are changed to losses of the same absolute magnitude. Kahneman and Tversky report evidence for this kind of effect from an experiment involving choices among prospects of the form $s_5 = (x, p; 0, 1 - p)$ and $r_5 = (y, \bullet p; 0, 1 - \bullet p)$. For given absolute values of x and y the majority of subjects revealed $s_5 > r_5$ when $y > x > 0$ and $r_5 > s_5$ when $y < x < 0$.

The "Asian disease" example discussed at the end of section 3 is consistent with the reflection effect. In that example, the choice between prospects was affected by the description of options. When outcomes were framed as lives saved, the majority of choosers were attracted to a sure gain of 200 out of 600 lives; when framed as losses the majority rejected the sure loss of 400 out of 600 deaths, preferring instead to take the risk. The effect observed there can be interpreted as a reflection effect with risk aversion in relation to gains and risk-seeking for losses. Before we could think this an explanation of the Asian disease problem, however, we need an account of how consequences are *interpreted*. From an objective standpoint, two hundred lives saved out of six hundred *is the same thing as* four hundred lives lost, hence a full explanation would require a theory of how framing affects whether an outcome is interpreted as a gain or a loss. Kahneman and Tversky go some way toward this in their discussion of editing.

Prospect theory assumes that prior to the second stage of evaluation, individuals will edit prospects using a variety of heuristics. One of the major editing operations involves the *coding* of outcomes as gains and losses relative to a reference point. Kahneman and Tversky argue that the reference point will typically be the current asset position, but they allow the possibility that "the location of the reference point, and the consequent coding of outcomes as gains or losses, can be affected by the formulation of the offered prospects, and by the expectations of the decision maker" (p. 274). Notice that this possibility of differential coding under the two problem descriptions is a necessary step in explaining responses to the Asian disease problem. While some economists might be tempted to think that questions about how reference points are determined sound more like psychological than economic issues, recent research is showing that understanding the role of references points may be an important step in explaining real economic behavior in the field (see, for example, Heath, Huddart, and Lang 1999).

Several of the other editing routines in prospect theory are essentially rules for simplifying prospects and transforming them into a form that can be more easily handled in the second phase. One such operation is the rule of *combination*, which simplifies prospects by combining the probabilities associated with identical outcomes. For example, a prospect *described* as $(x_1, p_1; x_1, p_2; x_3, p_3; \ldots)$ may be *evaluated* as the simplified prospect $(x_1, (p_1 + p_2); x_3, p_3; \ldots)$. Notice that these two prospects are not, in general, equivalent if $\bullet(\cdot)$ is nonlinear. Decision makers may also simplify prospects by rounding probabilities and/or outcomes. Further operations apply to *sets* of prospects. The operation of *cancellation* involves the elimination of elements common to the prospects under consideration.

Hence a choice between prospects $q' = (x, p; q, 1 - p)$ and $r' = (x, p; r, 1 - p)$ *may* be evaluated as a choice between q and r. Although cancellation is effectively an application of the independence axiom of EU, the editing phase does not imply that choices will generally satisfy independence, since whether a particular rule is applied depends upon whether or not it is salient. Although they have no formal theory of salience they do present evidence that editing is context dependent. One example shows that cancellation is used in some cases where it is salient and not in others (see their discussion of the "isolation effect," p. 271).

One further rule—I will call it the *dominance heuristic*—has the effect of eliminating stochastically dominated options from the choice set prior to evaluation. The addition of the dominance heuristic does not, however, remove all possibility of monotonicity violation. Kahneman and Tversky assume that individuals scan the set of options and delete dominated prospects *if they are detected*. This ensures the deletion of "transparently" dominated options but leaves open the possibility that some dominated options survive application of the routine. Since the preference function is not generally monotonic, such options may ultimately be chosen.

This strategy for imposing monotonicity has the further, perhaps surprising, implication that choices may be nontransitive. If • $(\cdot)$ is nonlinear, then prospect theory implies that there will be some q and r where q stochastically dominates r such that $V(r) > V(q)$.[18] So long as this dominance is transparent, the dominance heuristic ensures that there will be no *direct* violation of monotonicity and r will not be chosen over q. In general, however, it should be possible to find some other prospect s, such that $V(r) > V(s) > V(q)$. If there is no relation of dominance between s and either of q or r, then pairwise choice among these three gambles will generate a systematic cycle of choice in which $q >_c r$ and $r >_c s$ and $s >_c q$ where $>_c$ is the relation "is chosen over." Quiggin (1982, p. 327) calls this an "undesirable result."

Quiggin's reaction would not be untypical of economists more generally, most of whom have taken both transitivity and monotonicity to be fundamental principles that any satisfactory theory should embody. On the other hand, several economists, Quiggin included, have thought aspects of prospect theory appealing and have sought to build the relevant features into models more in keeping with conventional theoretical desiderata. For example, part of Quiggin's motivation in developing rank-dependent expected-utility theory was to establish that a central feature of prospect theory—nonlinear decision weights—can be built into a preference function without sacrificing monotonicity. By constructing decision weights cumulatively, we obtain a (transitive) preference function that is monotonic without the need for an additional editing routine. Papers by Starmer and Sugden (1989),

[18] To see how nonlinearity of $\pi(\cdot)$ can generate violations of monotonicity, consider a simple case where $q = (x, 1)$ and $r = (x - \varepsilon, p; x, 1 - p)$. Suppose $\varepsilon > 0$ hence q dominates r: If $\pi(\cdot)$ is concave, probabilities are overweighted, and the dominated option r is preferred for some ε. Now suppose $\varepsilon < 0$, hence r dominates q: if $\pi(\cdot)$ is convex, probabilities are underweighted, and the dominated option q is preferred for some ε.

Luce and Fishburn (1991), and Tversky and Kahneman (1992) show that the rank-dependent form can be extended to capture another key element of prospect theory: valuing outcomes relative to *reference points*.

In Starmer and Sugden's model, any prospect q is valued by the function $V(q) \bullet V^{+}(q) + V^{-}(q)$ where $V^{+}(q)$ is the rank-dependent expected utility of a *transformed* prospect q^{+}; this is equivalent to q excepting that any outcomes of q that are *losses* are replaced by zeros. Similarly, $V^{-}(q)$ is obtained by applying the standard rank-dependent form to a transformed prospect q^{-}, in this case, any outcomes that are *gains* are replaced by zeros. Tversky and Kahneman's model, *cumulative prospect theory*, is more general in that it allows the decision-weighting function to be different for the positive and negative components. The development of these so called sign- and rank-dependent models demonstrates that important aspects of prospect theory can be captured within a formal model that is essentially conventional, without the need to invoke an editing phase.

In these later models, the procedural element central to prospect theory has disappeared,[19] No doubt the abandonment of editing does leave some things unexplained. For instance, framing effects do suggest that choices are context-dependent in complex yet subtle ways, and the procedural approach seems to provide the more natural arena in which to model this. On the other hand, introducing elements of bounded rationality does considerably complicate the theoretical structure of models in ways that render them less compatible with the rest of economic theory. For example, working with a set of decision rules seems clumsy, relative to the neatness and tractability of optimizing a single function; unlike conventional models, procedural models often exhibit a degree of indeterminacy.[20]

Might such arguments provide sufficient grounds for defending a general theoretical presumption that agents behave "as if" fully rational? Conlisk (1996) reviews a series of methodological arguments that might be used to make such a case against incorporating ideas of bounded rationality into economics. He concludes that it is hard to make any convincing case against it. If that's correct, and I for one am persuaded, then the question to ask is whether departures from conventional models are of sufficient concern, from an empirical point of view, to justify the theoretical costs involved. I will say something about that in section 5.2, but first we consider an alternative avenue of departure from the conventional approach.

4.2.2. NONTRANSITIVE PREFERENCE THEORY

As we have seen, many have taken the view that the standard independence axiom of EU can be sacrificed for the sake of explaining the data. Transitivity, however, may be another matter. It might be tempting to think that transitivity is so fundamental to our ideas about preference that to give it up is to depart from theories of

[19] Although Tversky and Kahneman do mention that editing may be important, their 1992 model has no formal editing phase and their references to it are virtually asides.

[20] For instance, in prospect theory, the outcome of editing can depend on factors that are underdetermined by the theory, such as the order in which operations are applied (see Stevenson, Busmeyer, and Naylor 1991).

preference altogether. Can we speak of people maximizing anything if they don't have transitive preferences? It turns out that the answer is yes.

There is at least one well-known theory of choice based on a model of nontransitive preference. The theory I have in mind was proposed simultaneously by Bell (1982), Fishburn (1982), and Loomes and Sugden (1982). I will begin by discussing a version of this theory presented by Loomes and Sugden (1987), which they call *regret theory*. Its central premise is closely akin to the psychological intuition at the heart of the theory of disappointment. In that theory, it is assumed that an individual compares the outcomes *within a given prospect* giving rise to the possibility of disappointment when the outcome of a gamble compares unfavorably with what they might have had. Regret theory allows comparisons among consequences to affect choice, but in this case, the relevant comparisons occur among the consequences of alternative choice options.

Since the theory has to allow comparisons among choice options, it cannot be a conventional theory that assigns values independently to individual prospects. Loomes and Sugden propose a theory of pairwise choice in which preferences are defined over pairs of *acts*, where an act maps from *states of the world* to consequences.[21] Let *Ai* and *Aj* be two potential acts that result in outcomes x_{is} and x_{js}, respectively, in state of the world *S*. The utility of consequence x_{is} is given by a function $M(x_{is}, x_{js})$ which is increasing in its first argument and decreasing in its second. This function allows the utility from having x_{is} be suppressed by "regret" when $x_{is} < x_{js}$, or enhanced by "rejoicing" when $x_{is} > x_{js}$. The individual then seeks to maximize the expectation of modified utility $í_s\, p_s \cdot M(x_{is}, x_{js})$ where p_s is the probability of state *S*. Regret theory reduces to EU in the special case where $M(x_{is}, x_{js}) = u(x_{is})$.

Although preferences are defined over acts, the theory can be applied to choices among prospects given some assumption about how outcomes are correlated between them. One interesting case is when consequences are uncorrelated between prospects; that is, when prospects are *statistically independent*. In a choice between a pair of such prospects *q* and *r*, *if q* is chosen, the probability of getting x_i and missing out on x_j is given by $p_{qi} p_{rj}$ where p_{qi} is the probability of consequence x_i in *q* and p_{rj} the probability of x_j in prospect *r*. Preferences between *q* and *r* are then determined by the expression

$$q \underset{\prec}{\overset{\succ}{\sim}} r \Leftrightarrow \bullet_i \bullet_j\, p_{qi} p_{ij} \bullet \ (x_i, x_j) \underset{<}{\overset{>}{=}} 0, \tag{7}$$

where $\bullet\,(x_{is}, x_{js}) \bullet M(x_{is}, x_{js}) - M(x_{js}, x_{is})$. The function $\bullet\,(\cdot, \cdot)$ is skew symmetric by construction, hence $\bullet\,(x, y) \bullet - \bullet\,(y, x)$ and $\bullet(x, x) \bullet 0$ for all *x*, *y*.

If prospects are statistically independent, the addition of a further assumption, which Loomes and Sugden call *regrets-aversion*,[22] implies that indifference curves

[21] As a theory of pairwise choice, regret theory has limited applicability, but ways of generalizing the theory have been suggested by Sugden (1993) and Quiggin (1994).

[22] In their early discussions of regret theory, Loomes and Sugden called this assumption "convexity."

will fan out in the probability triangle. Regret-aversion requires that for any three consequences $x \succ y \succ z$, $\bullet(x, z) > \bullet(x, y) + \bullet(y, z)$. The interpretation of the assumption is that large differences between what you get from a chosen action and what you might have gotten from an alternative give rise to disproportionately large regrets; so people prefer greater certainty in the distribution of regret. Under these conditions, regret theory is equivalent to Chew and MacCrimmon's weighted-utility theory, and so indifference curves in the probability triangle will have the pattern described in figure 4.4 above (see Sugden [1986] for a simple demonstration of this). Consequently, regret theory is able to explain the standard violations of the independence axiom for statistically independent prospects.[23]

If we consider the class of all statistically independent prospects—not just those with up to three pure consequences—weighted-utility theory is a special case of regret theory. Specifically, the representation in expression 7 is obtained from Chew and MacCrimmon's axiom set by relaxing transitivity. This is the route by which Fishburn (1982) arrived at this model (he calls it *skew-symmetric bilinear utility* or SSB). Fishburn's model is identical with regret theory for statistically independent prospects, and we can think of regret theory as a generalization of SSB that extends it to nonindependent prospects: in this realm, regret-aversion has some very interesting implications.

Consider three stochastically equivalent actions A_1, A_2, and A_3, each of which gives each of the consequences $x \succ y \succ z$ in one of three equally probable states of the world s_1, s_2, and s_3. Any conventional theory entails a property of *equivalence*: that is, indifference among stochastically equivalent options, hence, for any such theory, $A_1 \sim A_2 \sim A_3$. In regret theory, however, it matters how consequences are assigned to states, and for particular assignments, regret theory implies a strict preference among stochastically equivalent acts, violating equivalence. For example, suppose that the three acts involved the following assignment of consequences to states:

	s_1	s_2	s_3
A_1	z	y	x
A_2	x	z	y
A_3	y	x	z

If we consider preferences between the first two acts, regret theory implies

$$A_1 \substack{\succ \\ \sim \\ \prec} A_2 \Leftrightarrow [\bullet(z,x) + \bullet(y,z) + \bullet(x,y) \substack{> \\ = \\ <} 0. \qquad (8)$$

[23] Some instances of the common consequence effect have involved statistically nonindependent options, and these cases are not consistent with regret theory (unless we assume agents treat options as if they are independent even when they are not).

Using the skew symmetry of • (·, ·), the term in square brackets is equal to [• (x, y) + • (y, z) − • (x, z)]. Assuming regret-aversion, this will be negative, hence regret theory implies a strict preference $A_2 \succ A_1$. It is easy to see that the same reasoning applied to the other two possible pairwise comparisons implies $A_3 \succ A_2$ and $A_1 \succ A_3$. Hence, regret theory also implies a *cycle of preference* of the form: $A_2 \succ A_1, A_3 \succ A_2, A_1 \succ x\, A_3$. Now consider adding some small positive amount • to one consequence of action A_1. The resulting action, call it A_1*, stochastically dominates each of the original actions. But since regret theory implies $A_2 \succ A_1$ we should expect $A_2 \succ A_1^*$ for at least some • > 0. Hence regret theory also implies violations of monotonicity.

Relative to the conventional approach then, preferences in regret theory are not at all well behaved: they satisfy neither monotonicity nor transitivity, and the theory allows strict preferences between stochastically equivalent acts. While such properties may seem peculiar to the eye of the conventional economist, from the descriptive angle, the crucial question is whether such implications of the theory are borne out by actual behavior. Shortly after proposing regret theory, Loomes and Sugden (1983) argued that at least one might be. Consider the following three acts labeled $, P, and M with monetary consequences $x \succ y \succ m \succ 0$ defined (for the sake of simplicity) over three equiprobable states:

	s_1	s_2	s_3
$	x	0	0
P	y	y	0
M	m	m	m

The actions labeled $ and P have the structure of typical $- and P-bets: they are binary gambles where $ has the higher prize, and P the higher probability of "winning"; the third act gives payoff m for sure. Loomes and Sugden show that, given regret-aversion, pairwise choices over acts with this structure may be cyclical, and if a cycle occurs, it will be in a specific direction with P ≻ $, M ≻ P and $ ≻ M. Now recall that in a standard experiment, subjects reveal P ≻ $ in a straight choice between options but place a higher value on $ relative to P in separate valuation tasks. If we interpret choices from {$, M} and {P, M} as analogues of valuation tasks asking "is $ (or P) worth more or less than m," then the cycle predicted by regret theory can be interpreted as a form of preference reversal.

So, regret theory offers the tantalizing opportunity of explaining violations of independence and preference reversal within a theory of preference maximization. Of course, since observation of preference reversal predates the development of regret theory, that phenomena offers only weak support for the unconventional predictive content of regret theory. More recent research has aimed at testing some novel predictions of regret theory and some of the results from this line of research are discussed in Starmer (2000).

5. Evaluating Alternatives to Expected-Utility Theory

5.1. The Recent Experimental Evidence

Starting in the mid-1980s, a number of researchers turned their attention toward testing nonexpected-utility theories. The majority of this work involved experimental testing, some of it designed to compare the predictive abilities of competing theories; some designed to test novel implications of particular theories; and some designed to test the descriptive validity of particular axioms. A very large volume of work has emerged in this arena, providing a much richer evidential base against which theories can be judged.

As we have seen, conventional theories all imply the existence of indifference curves in the probability triangle, and certain of their key properties can be expressed in terms of characteristics of the indifference maps they generate. For instance, Machina's theory implies generalized fanning-out, while other theories imply a mixture of fanning-in and fanning-out. A large number of experimental studies have explicitly examined individual behavior in choices among prospects in probability triangles. The data generated from these "triangle experiments" provides a vantage point from which we can ask the following question: suppose one were attempting to construct a conventional theory now, with the aim of accounting for the evidence currently available, are there any obvious properties one should seek to build in?

Although the evidence is both rich and complex, a number of stylized facts apply across a range of studies. In my view, three observations seem particularly robust. First, if you want a theory consistent with the available data *don't impose generalized fanning-out*. Evidence from a wide range of studies reveals behavior inconsistent with linear parallel indifference curves, but the patterns actually observed are more complex than generalized fanning-out. For example, while numerous studies reproduce behavior consistent with Allais paradox violations of EU in choice pairs moving left to right along the bottom edge of the probability triangle, another finding replicated across a range of studies—including Camerer (1989), Chew and Waller (1986), Battalio, Kagel, and Jiranyakul (1990), and Starmer (1992)—is a tendency for behavior to become less risk-averse moving up along the left-hand edge of probability triangles. Such behavior would be consistent with a tendency for indifference curves to fan in. These facts mitigate in favor of theories like disappointment-aversion, implicit utility, quadratic utility, and models with decision weights, all of which allow a mixture of fanning-in and fanning-out.

A second general lesson in the data seems to be *don't impose betweenness*. There is considerable evidence—a good part of it is reviewed in Camerer and Teck-Hua Ho (1994)—that choices are inconsistent with the assumption of linear indifference curves. Together these two requirements narrow the field considerably: if we want a theory of mixed fanning with nonlinear indifference curves, of the theories reviewed above the only contenders are quadratic utility, lottery-dependent utility, and models with decision weights.

A third widely observed finding arguably nudges the decision weighting models into the lead: *behavior on the interior of the probability triangle tends to conform more closely to the implications of EU than behavior at the borders*. Although significant off-border violations are observed in at least some experiments (see for example Wu and Gonzalez 1996), several studies, including those of Conlisk (1989); Camerer (1992); David Harless (1992); and Gigliotti and Sopher (1993) suggest that violations of EU are *concentrated* in comparisons between options involving prospects on or near the borders of triangles. It is important to note that this observation is unlikely to rescue EU for practical purposes. A natural interpretation of the "border effect" is that individuals are particularly sensitive to changes in the likelihood of outcomes with "extreme" probabilities (i.e., moving off the border of the triangle, we introduce a low probability event; in the vicinity of each corner, some outcome is near certain). It is very easy to think of important choice-scenarios involving real prospects with "extreme" probabilities; for example, individual decisions about participation in national or state lotteries or collective decisions about nuclear power generation involve high-magnitude outcomes (winning the lottery, suffering the effects of a radiation leak) occurring with very small probabilities. Consequently, there are good reasons to model sensitivity to "extreme" probabilities. One obvious way to do it is via decision weights.[24]

In summary, if one is looking to organize the data from the large number of triangle experiments, then the decision-weighting models are probably the best bet. Moreover, there is a striking degree of convergence across studies regarding the functional form to use; for best predictions the key ingredient seems to be an inverted s-shaped weighting function. Empirical support for this specification comes from a wide range of studies including Lattimore, Baker, and Witte (1992); Tversky and Kahneman (1992); Camerer and Ho (1994); Abdellaoui (1998); and Gonzalez and Wu (1999), all of which fit the decision-weighting model to experimental data. Collectively, these studies show that models with s-shaped probability transformations offer significant predictive improvement over EU and outperform other rivals. Most of the studies in this vein, at least those conducted in recent times, employ the rank-dependent transformation method, though different mathematical forms have been used for the probability-weighting function. Lattimore, Baker, and Witte use a probability weighting function of the form

$$\bullet\,(p_i) = \bullet p_i^{\bullet} \,/\, \left[\bullet p_i^{\bullet} \overset{n}{\underset{k=1}{+}} \bullet\, p_k^{\bullet}\right] \tag{9}$$

for $i, k = 1, 2, \ldots, n$, $k \bullet i$ and $\bullet, \bullet > 0$ (n is the number of outcomes as usual). This captures a number of other proposed forms (e.g., those of Uday Karmarkar 1978 and Quiggin 1982) as special cases. With $\bullet = \bullet = 1$, $\bullet(p_i) = p_i$, hence we get EU. More generally, the parameter $\bullet$ controls the inflection point and $\bullet < 1$ generates

[24] Another theoretical possibility suggested by Neilson (1992) is to allow the utility function defined over outcomes to depend on the number of outcomes: this generates different behavior on and off the border, but experimental tests of the model (see Stephen Humphrey 1998) have not been supportive.

the inverted-s with the consequent overweighting of "small" probabilities below the inflection point, and underweighting above it. With $\bullet < 1$, $\bullet(\cdot)$ is "sub-certain" in the sense that the sum of weights $(\bullet_i \bullet (p_i))$ will be less than unity. Lattimore, Baker, and Witte (1992, p. 381) describe this as " 'prospect pessimism' in the sense that the value of the prospect is reduced vis-à-vis certain outcomes." In their empirical estimates, they find that allowing nonlinear decision-weights offers significant improvement in predictive power over EU (which is the best model for only about 20 percent of their subjects). The best-fitting weighting function is generally the inverted-s exhibiting greater sensitivity to high and low probabilities relative to mid-range probabilities. They also report differences between the best-fitting weighting functions for gains and losses (for example "pessimism" is more pronounced for losses), though the interpretation of these differences is potentially confounded by the fact that, in their study, gains are measured in units of money while losses are measured in units of time.

Single-parameter weighting functions have been proposed by Tversky and Kahneman (1992) and Prelec (1998). Tversky and Kahneman suggest the form $\bullet(p) = p/[(p + (1 - p)^{\bullet})^{1/\bullet}]$. This generates the inverted-s for $0 < \bullet < 1$, and reducing $\bullet$ lowers the crossover point while accentuating the curvature of the function. Their empirical analysis supports the s-shaped weighting function and also reveals systematic differences in behavior for gains and losses: specifically, indifference curves in the best-fitting models for losses resemble those for gains flipped around a 45 degree line. This supports the case for a model that distinguishes between gains and losses (i.e., a model with a reference point), though virtually no work is done by the weighting function here; essentially, the same probability-weighting function works well for both gains and losses.

Prelec proposes the function $\bullet(p) = \exp(-(-ln\, p)^{\bullet})$. With $0 < \bullet < 1$, this generates the inverted-s with a fixed inflection point at $p = 1/e = 0.37$. Visually, $\bullet$ is the slope of $\bullet(\cdot)$ at the inflection point, and as $\bullet$ approaches unity, $\bullet(\cdot)$ becomes approximately linear; as it approaches zero, $\bullet(\cdot)$ approximates a step function. Prelec argues that a crossover in the vicinity of $1/e$ is consistent with the data observed across a range of studies. A novel feature of Prelec's contribution is to provide an axiomatization for this form, and he also discusses a two-parameter generalization. The two-parameter version is similar in spirit to the "linear in log odds form" proposed by Gonzalez and Wu (1999) in that it allows the curvature and elevation of the weighting function to be manipulated (more or less) independently. In the latter form, probability weights are given by

$$\bullet(p_i) = \bullet\, p_i^{\bullet} / [\bullet\, p_j^{\bullet} + (1 - p_i)^{\bullet}]. \tag{10}$$

The parameter $\bullet$ primarily controls the absolute value of $\bullet(\cdot)$ by altering the elevation of the function, relative to the 45-degree line, while $\bullet$ primarily controls curvature. Gonzalez and Wu's data suggests that the flexibility of a two-parameter model may be useful for explaining differences among individuals. For other purposes, however, parsimony favors the one-parameter versions.

Conventional theory can claim a success here: a one-parameter extension to EU can offer significantly improved predictive power for a large body of data generated mainly from triangle experiments. If we want to predict behavior over simple choices like this, we know a good deal about how to improve on EU. Notwithstanding this success, it is important to note that there is a wide range of evidence that conventional theories stand little chance of digesting. For example, there is considerable evidence revealing systematic failures of monotonicity and transitivity in risky choice experiments. Some of this evidence is reviewed in Starmer (2000).

5.2. Evidence from the Field

I have heard some economists argue that they would take more notice of non-EU models if they could be shown cases where they help to explain real-world phenomena of practical interest to economics. It is a fair point, but proponents of nonexpected utility theory can muster some strong responses. Let me illustrate this by way of a couple of examples.

The standard theory of insurance based on EU has some implications that have long been regarded as highly implausible. For example, a risk-averse expected-utility maximizer will not buy full insurance in the presence of positive marginal loading (see Mossin 1968). This implication, Karl Borch (1974) suggests, is "against all observation." More recently, Wakker, Thaler, and Tversky (1997) have made a similar point in relation to "probabilistic insurance." Think of probabilistic insurance as a policy with some fixed probability q that a claim will not be paid in the event of an insured loss. Wakker, Thaler, and Tversky show that an expected-utility maximizer willing to pay a premium c for full insurance against some risk should be willing to pay a premium approximately equal to the actuarially adjusted premium $(1 - q) \cdot c$ for probabilistic insurance. Survey evidence, however, shows that people are extremely averse to probabilistic insurance and their willingness to pay for it is much less than standard theory allows.

If expected utility can't explain insurance behavior, can nonexpected-utility theory do any better? Part of the answer is provided by Segal and Spivak (1990), who show that a number of implications of EU for insurance and asset demand that are widely recognized to be counterintuitive have a common origin. They arise because, with any smooth (i.e., differentiable) utility function, EU implies that agents will be approximately risk neutral for small risks (since the utility function will be almost linear). This theoretical property is at odds with peoples' actual risk attitudes as revealed through their reactions to probabilistic insurance and so on: people demand a much greater reduction in premium than the actuarially fair adjustment for accepting a small positive risk of claim nonpayment.

Segal and Spivak go on to show that the counterintuitive implications of EU carry through to nonexpected-utility theories which have similar smoothness properties. This captures a large number of alternatives to EU and, in fact, only a single type of theory escapes their net: the decision-weighting models. It is easy to see why models with probability transformations do not imply approximate risk neutrality for small risks since risk averse behavior can be generated by nonlinear probability weighting even where the utility function is linear. So, for example, aversion to probabilistic insurance is easily explained by overweighting of the small probability of nonpayment. As such, decision-weighting models stand out as leading contenders to explain aspects of insurance behavior that it has long been known standard theory cannot handle. There is growing evidence that probability weighting may be an important ingredient in explaining a variety of field data relating to gambling and insurance behavior and several examples are discussed by Camerer (2000).

Another field phenomenon that has perplexed economists is the size and persistence of the excess return on stocks over fixed income securities. This is the so called *equity premium puzzle* and it is the economics equivalent of the crop circle: we have seen it in the field, but we have real trouble explaining how it got there. Since the return on stocks is more variable, standard theory is consistent with some difference in the long-run rates of return, but since Mehra and Prescott (1985) it has been recognized that the observed disparity implies implausibly high degrees of risk-aversion in standard models of asset pricing. One possible explanation for (part of) the equity premium has been suggested by Epstein and Zin (1990). They show that a recursive utility model using rank-dependent preferences predicts an equity premium, though only about one third of the size that is usually observed. A full, and in my view much more convincing, account has been suggested by Benartzi and Thaler (1995) who show that the level of equity premium is consistent with prospect theory, with the added assumption that agents are myopic (i.e., they assess expected returns over "short" time horizons). The crucial element of prospect theory for this explanation is loss-aversion. In the short run, there is a significant chance that the return to stocks is negative so if, as loss-aversion implies, investors are particularly sensitive to these possible negative returns, that would explain the equity premium for myopic investors. But just how loss-averse and how myopic do agents have to be for this explanation to work? Benartzi and Thaler show that, assuming people are roughly twice as sensitive to small losses as to corresponding gains (which is broadly in line with experimental data relating to loss-aversion), the observed equity premium is consistent with the hypothesis that investments are evaluated annually. This is a very simple, and to my mind, intuitively appealing account of another important field phenomenon which has defied explanation in standard theory.

Notice that while loss-aversion can be accommodated in conventional models like the sign- and rank-dependent theories, the other ingredient in this explanation of the equity premium—i.e., myopia—belongs in another tradition. This is essentially a bounded rationality assumption, and while the one-year time horizon has a nice ring of plausibility to it, it sits much more naturally alongside procedural theories like the original version of prospect theory. Bounded rationality assumptions seem to be providing the missing links necessary to explain an increasing range of economic phenomena (see Camerer 1998 for a recent review of applications in individual decision making).

5.3. Theoretical Applications

While a good deal of effort has been devoted to developing alternatives to EU, by comparison, the use of such models in theoretical work outside of the specialist literature has been limited. Does this suggest that alternative models are too complex or intractable to be useful in a broader theoretical context? In general, I think the answer is no and that other factors, including the sheer variety of alternatives, most likely explain the relatively slow take-up of new models.

Although EU has been a central building block in core areas of economics, many tools and results that have been developed using it actually require weaker assumptions (see for example, Machina 1982, 1987, 1989a; Karni and Safra 1989, 1990; Crawford 1990). That said, it is true that giving up EU has dramatic implications in some areas of theory, and one pertinent example is the area of dynamic choice. If EU does not hold, then sequential choices may be *dynamically inconsistent*. To appreciate the significance of this, consider a sequential choice problem represented by a standard decision tree. An agent who is dynamically inconsistent may identify an optimal path viewed from the initial choice node, but then be unwilling to take actions that form part of that optimal path at choice nodes further down the tree. Wakker (1999) suggests an analogy between dynamic inconsistency and schizophrenia: the dynamically inconsistent agent has something akin to a split personality, with different aspects of the person revealing themselves in different parts of the tree. Although some might regard this as a "problem" with nonexpected utility models, I think that this conclusion could be misleading for two reasons, one theoretical, the other empirical.

From the theoretical point of view, it is important to note that relaxation of independence does not necessarily imply dynamic inconsistency. Machina (1989b) has shown that agents with nonexpected-utility preferences can be dynamically consistent if we are prepared to sacrifice the assumption of *consequentialism*. An implication of consequentialism in standard decision-tree analysis is that agents are entirely forward looking: at any given decision node, the consequentialist decision maker ignores any part of the tree that cannot be reached moving forward from that node. In contrast, Machina argues that risks borne in the past may be relevant to current decisions and he provides some telling examples of where that could be the case. As such he defends the notion of a dynamically consistent non-EU agent by rejecting consequentialism.

It has only recently been properly understood that axioms of EU, including the independence axiom, follow from assuming certain principles of dynamic choice (see Hammond 1988; McClennen 1990; Cubitt 1996). This provides a new form of normative defence for EU. On the other hand, since we know that independence fails empirically, at least one of the dynamic choice principles that jointly imply it must be failing too. It follows that if we want to predict the behavior of real agents in dynamic contexts, we will need models of dynamic decision making that relax the suspect dynamic choice principle(s) implicit in EU.

It has to be said that, overall, the volume of work applying nonexpected-utility models looks quite small given how long some of the theories have been available. I think that things may be changing and that we will see increasing use of models based on the rank-dependent form. Until recently, the sheer variety of competing models probably counted against their use. Too many alternatives were on offer with no obvious way to discriminate among them (bear in mind that many of these theories were proposed to explain the same, relatively small, set of choice anomalies). But now that much more evidence has accumulated, it seems clear that there are quantitatively important phenomena that should not be ignored in general economic analysis. One of these is surely the phenomenon of nonlinear probability weighting. The rank-dependent model is likely to become more widely used precisely because it captures this robust empirical phenomenon in a model that is quite amenable to application within the framework of conventional economic analysis.

Loss-aversion is another empirically important concept, and I sense that economists are becoming more interested in studying the implications of assuming loss-averse preferences for a range of economic issues. Tversky and Kahneman (1991) present a model—based on prospect theory—that applies the ideas of reference dependence and loss-aversion in *riskless* choice, and attempts are currently underway to examine the implications of rank-dependent preferences for fundamental theoretical issues in economics. For example, Munro (1998) examines the implications for welfare economics of assuming reference-dependent preferences; Munro and Sugden (1998) examine the conditions necessary for general equilibrium in an economy where agents have reference-dependent preferences.

Sign- and rank-dependent models—like cumulative prospect theory—capture both of these empirically important phenomena in a theoretically compact way. And, while not all of the empirical evidence fits this approach, it does provide an account consistent with some of the most robust stylized facts from a range of experimental studies.[25] Since these models are essentially conventional, and since their use seems to be expanding, general claims to the effect that they are intractable, or not useful in economics more broadly, seem unconvincing.

Perhaps there is a case for thinking that the position we should now aim for is one in which models like cumulative prospect theory become the default in applied economics with EU used as a convenient special case, but only when we can be confident that loss-aversion and probability weighting are insignificant. While that position may be some way off, my prediction is that the use of models incorporating probability weights and loss-aversion will grow rapidly, and my normative judgment is that, if it doesn't, it ought to.

References

Abdellaoui, Mohammed. 1998. "Parameter-Free Eliciting of Utilities and Probability Weighting Functions." Working paper, Groupe de Recherche sur le Risquel' Information et la Décìsìon (GRID), École Normale Supérieure (ENS), Cachan, France.

———. 1999. "A Genuine Rank-Dependent Generalization of von Neumann-Morgenstern Expected Utility Theorem." Working paper, GRID, ENS, Cachan, France.

[25] For those interested in where rank-dependent models fail, see Wakker (1994), Wu (1994), and Birnbaum, Patton, and Lott (1999).

Ali, Mukhtar M. 1977. “Probability and Utility Estimates for Racetrack Bettors.” *Journal of Political Economics*, 85: 803–15.
Allais, Maurice. 1953. “Le Comportement de l’Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l’Ecole Americaine.” *Econometrica*, 21: 503–46.
———. 1979. “The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School.” In *Expected Utility Hypotheses and the Allais Paradox*, edited by M. Allais and O. Hagen. Dordrecht: Reidel.
Anand, Paul. 1993. *Foundations of Rational Choice under Risk*. Oxford: Oxford University Press and Clarendon.
Battalio, Raymond C., John. H. Kagel, and Komain Jiranyakul. 1990. “Testing between Alternative Models of Choice under Uncertainty: Some Initial Results.” *Journal of Risk Uncertainty*, 3: 25–50.
Becker, Joao L., and Rakesh K. Sarin. 1987. “Gamble Dependent Utility.” *Management Science*, 33: 1367–82.
Bell, David. 1982. “Regret in Decision Making under Uncertainty.” *Operations Research*, 20: 961–81.
———. 1985. “Disappointment in Decision Making under Uncertainty.” *Operations Research*, 33: 1–27.
Benartzi, Shlomo, and Richard H. Thaler. 1995. “Myopic Loss Aversion and the Equity Premium Puzzle.” *Quarterly Journal of Economics*, 110: 73–92.
Bernoulli, Daniel. 1954 [1738]. “Exposition of a New Theory on the Measurement of Risk.” *Econometrica*, 22: 23–26.
Birnbaum, Michael H., Jamie N. Patton, and Melissa K. Lott. 1999. “Evidence against Rank-Dependent Utility Theories: Tests of Cumulative Independence, Interval Independence, Stochastic Dominance, and Transitivity.” *Organizational Behavior and Human Decision Processes*, 77: 44–83.
Borch, Karl H. 1974. *The Mathematical Theory of Insurance*. Lexington, MA: Lexington Books.
Camerer, Colin F. 1989. “An Experimental Test of Several Generalised Utility Theories.” *Journal of Risk Uncertainty*, 2(1): 61–104.
———. 1992. “Recent Tests of Generalizations of Expected Utility Theory.” In *Utility: Theories Measurement*, *and Applications*, edited by Ward Edwards. Norwell, MA: Kluwer.
———. 1995. “Individual Decision Making.” In *Handbook of Experimental Economics*, edited by J. Kagel and A. E. Roth. Princeton: Princeton University Press.
———. 1998. “Bounded Rationality in Individual Decision Making.” *Experimental Economics*, 1: 163–83.
———. 2000. “Prospect Theory in the Wild: Evidence from the Field.” In *Choices, Values and Frames*, edited by Daniel Kahneman and Amos Tversky. Cambridge: Cambridge University Press.
Camerer, Colin F., and Teck-Hua Ho. 1994. “Violations of the Betweenness Axiom and Nonlinearity in Probability.” *Journal of Risk Uncertainty*. 8: 167–96.
Chateauneuf, Alain, and Michèle Cohen. 1994. “Risk-seeking with Diminishing Marginal Utility in a Non-expected Utility Model.” *Journal of Risk and Uncertainty*, 9: 77–91.
Chew, Soo Hong. 1983. “A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox.” *Econometrica*, 51: 1065–92.
———. 1989. “Axiomatic Utility Theories with the Betweenness Property.” *Annals Operational Research*, 19: 273–98.

Chew, Soo Hong, and Larry G. Epstein. 1989. "A Unifying Approach to Axiomatic Non-expected Utility Theories." *Journal of Economic Theory*, 49: 207–40.

Chew, Soo Hong, Larry G. Epstein, and Uzi Segal. 1991. "Mixture Symmetry and Quadratic Utility." *Econometrica*, 59: 139–63.

Chew, Soo Hong, Larry G. Epstein, and Peter P. Wakker. 1993. "A Unifying Approach to Axiomatic Non-Expected Utility Theories: Correction and Comment." *Journal of Economic Theory*, 59: 183–88.

Chew, Soo Hong, Edi Karni, and Zvi Safra. 1987. "Risk Aversion in the Theory of Expected Utility with Rank-dependent Probabilities." *Journal of Economic Theory*, 42: 370–81.

Chew, Soo Hong, and Kenneth MacCrimmon. 1979a. "Alpha-nu Choice Theory: A Generalisation of Expected Utility Theory." Working paper 669, University of British Columbia.

———. 1979b. "Alpha Utility Theory, Lottery Composition, and the Allais Paradox." Faculty of Commerce and Business Admin. Working paper 686, University of British Columbia.

Chew, Soo Hong, and William S. Waller. 1986. "Empirical Tests of Weighted Utility Theory." *Journal of Mathematics and Psychology*, 30: 55–72.

Conlisk, John. 1989. "Three Variants on the Allais Example." *American Economic Review*, 79: 392–407.

———. 1996. "Why Bounded Rationality." *Journal of Economic Literature*, 34(2): 669–700.

Crawford, Vincent P. 1990. "Equilibrium without Independence." *Journal of Economic Theory*, 50: 127–54.

Cubitt, Robin P. 1996. "Rational Dynamic Choice and Expected Utility Theory." *Oxford Economical Papers*, 48: 1–19.

Dekel, Eddie. 1986. "An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom." *Journal of Economic Theory*, 40: 304–18.

Diecidue, Enrico, and Peter P. Wakker. 1999. "On the Intuition of Rank-dependent Expected Utility." Paper presented at FUR IX, Marrakesh.

Edwards, Ward. 1955. "The Prediction of Decisions among Bets." *Journal of Experimental Psychology*, 50: 201–14.

———. 1962. "Subjective Probabilities Inferred from Decisions." *Psychology Review*, 69: 109–35.

Epstein, Larry G., and Stanley E. Zin. 1990. " 'First-order' Risk Aversion and the Equity Premium Puzzle." *Journal of Monetary Economics*, 26: 387–407.

Fishburn, Peter C. 1978. "On Handa's 'New Theory of Cardinal Utility' and the Maximization of Expected Return." *Journal of Political Economy*, 86: 321–24.

———. 1982. "Nontransitive Measurable Utility." *Journal of Mathematics and Psychology*, 26: 31–67.

———. 1983. "Transitive Measurable Utility." *Journal of Economic Theory*, 31: 293–317.

———. 1988. *Nonlinear Preference and Utility Theory*. Brighton: Wheatsheaf Books.

Gigliotti, Garry, and Barry Sopher. 1993. "A Test of Generalized Expected Utility Theory." *Theory and Decision*, 35: 75–106.

Gonzalez, Richard, and George Wu. 1999. "On the Shape of the Probability Weighting Function." *Cognitive Psychology*, 38: 129–66.

Green, Jerry, and Bruno Jullien. 1988. "Ordinal Independence in Nonlinear Utility Theory." *Journal of Risk Uncertainty*, 1: 355–87.

Gul, Faruk. 1991. "A Theory of Disappointment in Decision Making under Uncertainty." *Econometrica*, 59: 667–86.

Hagen, Ole. 1979. "Towards a Positive Theory of Preferences under Risk." In *Expected Utility Hypotheses and the Allais Paradox*, edited by M. Allais and O. Hagen. Dordrecht: Reidel.

Hammond, Peter. 1988. “Consequentialist Foundations for Expected Utility.” *Theory and Decision*, 25: 25–78.
Handa, Jagdish. 1977. “Risk, Probability, and a New Theory of Cardinal Utility.” *Journal of Political Economy*, 85: 97–122.
Harless, David W. 1992. “Predictions about Indifference Curves inside the Unit Triangle: A Test of Variants of Expected Utility.” *Journal of Economic Behavior Organization*, 18: 391–414.
Hausman, Daniel. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.
Heath, Chip, Steven Huddart, and Mark Lang. 1999. “Psychological Factors and Stock Option Exercise.” *Quarterly Journal Economics*, 114(2): 601–27.
Holt, Charles A. 1986. “Preference Reversals and the Independence Axiom.” *American Economic Review*, 76: 508–15.
Humphrey, Stephen J. 1995. “More Mixed Results on Boundary Effects.” *Economic Letters*, 61: 79–84.
Jullien, Bruno, and Bernard Salanié. 1997. “Estimating Preferences under Risk: The Case of Racetrack Bettors.” Working paper, IDEI and GREMAQ, Toulouse University.
Kahneman, Daniel and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica*, 47(2): 263–91.
Karmarkar, Uday S. 1978. “Subjectively Weighted Utility: A Descriptive Extension of the Expected Utility Model.” *Org. Behavior Human Performance*, 21: 61–72.
Karni, Edi, and Zvi Safra. 1987. “ ‘Preference Reversal’ and the Observability of Preferences by Experimental Methods.” *Econometrica*, 55: 675–85.
———. 1989. “Dynamic Consistency, Revelations in Auctions and the Structure of Preferences.” *Review of Economic Studies*, 56: 421–34.
———. 1990. “Behaviorally Consistent Optimal Stopping Rules.” *Journal of Economic Theory*, 51: 391–402.
Kroll, Yoram, Haim Levy, and Amnon Rapoport. 1988. “Experimental Tests of the Separation Theorem and the Capital Asset Pricing Model.” *American Economic Review*, 78: 500–19.
Lattimore, Pamela K., Joanna R. Baker, and Ann D. Witte. 1992. “The Influence of Probability on Risky Choice: A Parametric Examination.” *Journal of Economic Behavior Organization*, 17: 377–400.
Lavoie, Marc. 1992. *Foundations of Post-Keynesian Economic Analysis*. Aldershot, UK: Edward Elgar.
Leland, Jonathan W. 1998. “Similarity Judgements in Choice under Uncertainty: A Reinterpretation of the Predictions of Regret Theory.” *Management Science*, 44: 659–72.
Lichtenstein, Sarah, and Paul Slovic. 1971. “Reversals of Preference between Bids and Choices in Gambling Decisions.” *Journal of Experimental Psychology*, 89: 46–55.
Lindman, Harold R. 1971. “Inconsistent Preferences among Gambles.” *Journal of Experimental Psychology*, 89: 390–97.
Loomes, Graham, Peter Moffat, and Robert Sugden. 1998. “A Microeconometric Test of Alternative Stochastic Theories of Risk Choice.” ERC discussion paper 9806, University of East Anglia.
Loomes, Graham, and Robert Sugden. 1982. “Regret Theory: An Alternative Theory of Rational Choice under Uncertainty.” *Economics Journal*, 92: 805–24.
———. 1983. “A Rationale for Preference Reversal.” *American Economic Review*, 73: 428–32.
———. 1986. “Disappointment and Dynamic Consistency in Choice under Uncertainty.” *Review of Economic Studies*, 53(2): 271–82.

———. 1987. "Some Implications of a More General Form of Regret Theory." *Journal of Economic Theory*, 41(2): 270–87.

Luce, R. Duncan, and Peter C. Fishburn. 1991. "Rank and Sign-dependent Linear Utility Models for Finite First-Order Gambles." *Journal of Risk Uncertainty*, 4: 29–59.

MacCrimmon, Kenneth, and Stig Larsson. 1979. "Utility Theory: Axioms versus Paradoxes." In *Expected Utility Hypotheses and the Allais Paradox*, edited by M. Allais and O. Hagen. Dordrecht: Reidel.

Machina, Mark J. 1982. " 'Expected Utility' Theory without the Independence Axiom." *Econometrica*, 50: 277–323.

———. 1983. "The Economic Theory of Individual Behavior toward Risk: Theory, Evidence, and New Directions." Technical Report 433, Department of Economics, Stanford University.

———. 1987. "Choice under Uncertainty: Problems Solved and Unsolved." *Journal of Economic Perspectives*, 1: 121–54.

———. 1989a. "Comparative Statics and Non-Expected Utility Preferences." *Journal of Economic Theory*, 47: 393–405.

———. 1989b. "Dynamic Consistency and Non-Expected Utility Models of Choice under Uncertainty." *Journal of Economic Literature*, 27: 1622–68.

———. 1994. "Review of 'Generalized Expected Utility Theory: The Rank-Dependent Model' " *Journal of Economic Literature*, 32(3): 1237–38.

Marschak, Jacob. 1950. "Rational Behavior, Uncertain Prospects, and Measurable Utility." *Econometrica*, 18: 111–41.

McClennen, Edward F. 1990. *Rationality and Dynamic Choice: Foundational Explorations*. Cambridge: Cambridge University Press.

Mehra, Rajneesh, and Edward C. Prescott. 1985. "The Equity Premium Puzzle." *Journal of Monetary Economics*, 15: 145–61.

Moskowitz, Herbert. 1974. "Effects of Problem Presentation and Feedback on Rational Behavior in Allais and Morlat-Type Problems." *Decision Sciences*, 5: 225–42.

Mossin, Jan. 1968. "Aspects of Rational Insurance Purchasing." *Journal of Political Economy*, 76: 553–68.

Munro, Alistair. 1998. "Loss Aversion and Welfare Economics." Economics Research Centre working paper, University of East Anglia.

Munro, Alistair, and Robert Sugden. 1998. "A Theory of General Equilibrium with Reference Dependent Preferences." Mimeo, University East Anglia.

Neilson, William S. 1992. "A Mixed Fan Hypothesis and its Implications for Behavior toward Risk." *Journal of Economic Behavior Organization*, 19: 197–211.

———. 1979. "Some Mixed Results on Boundary Effects." *Economic Letters*, 39: 275–78.

Neumann, John von, and Oskar Morgenstern. 1947. *The Theory of Games and Economic Behavior*. 2nd ed. Princeton: Princeton University Press.

Payne, John W., James R. Bettman, and Eric J. Johnson. 1993. *The Adaptive Decision Maker*. Cambridge: Cambridge University Press.

Pidgeon, Nick, C. Hood, D. Jones, B. Turner, and R. Gibson. 1992. "Risk Perception." In *Risk: Analysis*, *Perceptions and Management*. Report of a Royal Society Study Group. London: The Royal Society.

Prelec, Drazen. 1998. "The Probability Weighting Function." *Econometrica*, 66: 497–527.

Preston, Malcolm G., and Phillip Baratta. 1948. "An Experimental Study of the Auction-Value of an Uncertain Outcome." *American Journal of Psychology*, 61: 183–93.

Quiggin, John. 1982. "A Theory of Anticipated Utility." *Journal of Economic Behavior Organization*, 3(4): 323–43.

———. 1994. "Regret Theory with General Choice Sets." *Journal of Risk Uncertainty*, 8(2): 153–65.
Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica*, 68(5): 1281–92.
Rubinstein, Ariel. 1988. "Similarity and Decision-Making Under Risk." *Journal of Economic Theory*, 46: 145–53.
Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: Wiley.
Schoemaker, Paul. 1982. "The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations." *Journal of Economic Literature*, 20(2): 529–63.
Segal, Uzi. 1988. "Does the Preference Reversal Phenomenon Necessarily Contradict the Independence Axiom?" *American Economic Review*, 78: 233–36.
———. 1989. "Anticipated Utility: A Measure Representation Approach." *Annals Operational Research*, 19: 359–73.
———. 1990. "Two-Stage Lotteries without the Independence Axiom." *Econometrica*, 58: 349–77.
———. 1993. "The Measure Representation: A Correction." *Journal of Risk Uncertainty*, 6: 99–107.
———. 1997. "Dynamic Consistency and Reference Points." *Journal of Economic Theory*, 72: 208–19.
Segal, Uzi, and Avia Spivak. 1990. "First Order versus Second Order Risk Aversion." *Journal of Economic Theory*, 51: 111–25.
Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics*, 69: 99–118.
Slovic, Paul. 1995. "The Construction of Preferences." *American Psychologist*, 50: 364–71.
Slovic, Paul, and Amos Tversky. 1974. "Who Accepts Savage's Axiom?" *Behavioral Science*, 19: 368–73.
Starmer, Chris. 1992. "Testing New Theories of Choice under Uncertainty Using the Common Consequence Effect." *Review of Economic Studies*, 59: 813–30.
———. 2000. "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk." *Journal of Economic Literature*, 38: 332–82.
Starmer, Chris, and Robert Sugden. 1989. "Violations of the Independence Axiom in Common Ratio Problems: An Experimental Test of Some Competing Hypotheses." *Annals of Operational Research*, 19: 79–102.
Stevenson, Mary K., Jerome R. Busmeyer, and James C. Naylor. 1991. "Judgement and Decision-Making Theory." In *New Handbook of Industrial Organizational Psychology*, edited by M. Dunnette and L. N. Hough. Palo Alto, CA: Consulting Psychology Press.
Sugden, Robert. 1986. "New Developments in the Theory of Choice under Uncertainty." *Bulletin of Economic Research*, 38: 1–24.
———1991. "Rational Choice: A Survey of Contributions from Economics and Philosophy." *Economic Journal*, 101: 751–85.
———. 1993. "An Axiomatic Foundation for Regret Theory." *Journal of Economic Theory*, 60: 159–80.
———. In Press. "Alternatives to Expected Utility Theory: Foundations and Concepts." In *Handbook of Utility Theory*, edited by P. J. Hammond and C. Seidl. Dordrecht: Kluwer.
Tammi, Timo. 1997. "Essays on the Rationality of Experimentation in Economics: The Case of Preference Reversal." Ph.D. diss. University of Joensuu, Finland.
Thaler, Richard H., and William T. Ziemba. 1988. "Parimutuel Betting Markets: Racetracks and Lotteries." *Journal of Economic Perspectives*, 2: 161–74.

Tversky, Amos. 1969. "Intransitivity of Preferences." *Psychology Review*, 76: 31–48.

Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science*, 211: 453–58.

———. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quarterly Journal of Economics*, 106(4): 1039–61.

———. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk Uncertainty*, 5(4): 297–323.

Tversky, Amos, Paul Slovic, and Daniel Kahneman. 1990. "The Causes of Preference Reversal." *American Economic Review*, 80(1): 204–17.

Tversky, Amos, and Richard H. Thaler. 1990. "Preference Reversals." *Journal of Economic Perspectives*, 4: 201–11.

Tversky, Amos, and Peter P. Wakker. 1995. "Risk Attitudes and Decision Weights." *Econometrica*, 63: 1255–80.

Wakker, Peter P. 1994. "Separating Marginal Utility and Probabilistic Risk Aversion." *Theory and Decision*, 36: 1–44.

———. 1999. "Justifying Bayesianism by Dynamic Decision Principles." Plenary paper presented at FUR IX, Marrakesh.

Wakker, Peter P., Ido Erev, and Elke U. Weber. 1994. "Comonotonic Independence: The Critical Test between Classical and Rank-Dependent Utility Theories." *Journal of Risk Uncertainty*, 9: 195–230.

Wakker, Peter P., Richard H. Thaler, and Amos Tversky. 1997. "Probabilistic Insurance." *Journal of Risk Uncertainty*, 15: 7–28.

Wu, George. 1994. "An Empirical Test of Ordinal Independence." *Journal of Risk Uncertainty*, 9: 39–60.

Wu, George, and Richard Gonzalez. 1996. "Curvature of the Probability Weighting Function." *Management Science*, 42: 1676–90.

Yaari, Menahem E. 1987. "The Dual Theory of Choice Under Risk." *Econometrica*, 55: 95–115.

CHAPTER 5

Prospect Theory in the Wild: Evidence from the Field

COLIN F. CAMERER

THE WORKHORSES of economic analysis are simple formal models that can explain naturally occurring phenomena. Reflecting this taste, economists often say they will incorporate more psychological ideas into economics if those ideas can parsimoniously account for field data better than standard theories do. Taking this statement seriously, this article describes ten regularities in naturally occurring data that are anomalies for expected utility theory but can all be explained by three simple elements of prospect theory: loss-aversion, reflection effects, and nonlinear weighting of probability; moreover, the assumption is made that people isolate decisions (or edit them) from others they might be grouped with (Read, Loewenstein, and Rabin 1999; cf. Thaler 1999). I hope to show how much success has already been had applying prospect theory to field data and to inspire economists and psychologists to spend more time in the wild.

The ten patterns are summarized in table 5.1. To keep the article brief, I sketch expected-utility and prospect theory very quickly. (Readers who want to know more should look elsewhere in this volume or in Camerer 1995 or Rabin 1998). In expected utility, gambles that yield risky outcomes x_i with probabilities p_i are valued according to $\Sigma\, p_i\, u(x_i)$, where $u(x)$ is the *utility* of outcome x. In prospect theory they are valued by $\Sigma\pi(p_i)v(x_i - r)$, where $\pi(p)$ is a function that weights probabilities nonlinearly, overweighting probabilities below .3 or so and underweighting larger probabilities.[1] The value function $v(x - r)$ exhibits diminishing marginal sensitivity to deviations from the reference point r, creating a "reflection effect" because $v(x - r)$ is convex for losses and concave for gains (i.e., $v''(x - r) > 0$ for $x < r$ and $v''(x - r) < 0$ for $x > r$). The value function also exhibits *loss aversion* if the value of a loss $-x$ is larger in magnitude than the value of an equal-sized gain (i.e., $-v(-x) > v(x)$ for $x > 0$).

The research was supported by NSF grant SBR-9601236 and the hospitality of the Center for Advanced Study in Behavioral Sciences during 1997–98. Linda Babcock and Barbara Mellers gave helpful suggestions.

[1] In rank-dependent approaches, the weights attached to outcomes are differences in weighted cumulative probabilities. For example, if the outcomes are ordered $x_1 > x_2 > \cdots > x_n$, the weight on outcome x_i is $\pi(p_i + p_2 + \cdots + p_i) - \pi(p_1 + p_2 + \cdots + p_{i-1})$. (Notice that if $\pi(p) = p$ this weight is just the probability p_i). In cumulative prospect theory, gains and losses are ranked and weighted separately (by magnitude).

Table 5.1
Ten Field Phenomena Inconsistent with EU and Consistent with Cumulative Prospect Theory

Domain	*Phenomenon*	*Description*	*Type of Data*	*Isolated Decision*	*Ingredients*	*References*
Stock market	**Equity premium**	Stock returns are too high relative to bond returns	NYSE stock, bond returns	Single yearly return (not long-run)	Loss-aversion	Benartzi and Thaler (1995)
Stock market	**Disposition effect**	Hold losing stocks too long, sell winners too early	Individual investor trades	Single stock (not portfolio)	Reflection effect	Odean (in press), Genesove and Mayer (in press)
Labor economics	**Downward-sloping labor supply**	NYC cabdrivers quit around daily income target	Cabdriver hours, earnings	Single day (not week or month)	Loss-aversion	Camerer et al. (1997)
Consumer goods	**Asymmetric price elasticities**	Purchases more sensitive to price increases than to cuts	Product purchases (scanner data)	Single product (not shopping cart)	Loss-aversion	Hardie, Johnson, Fader (1993)
Macroeconomics	**Insensitivity to bad income news**	Consumers do not cut consumption after bad income news	Teachers' earnings, savings	Single year	Loss-aversion, reflection effect	Shea (1995); Bowman, Minehar, and Rabin (1999)
Consumer choice	**Status quo bias, Default bias**	Consumers do not switch health plans, choose default insurance	Health plan, insurance choices	Single choice	Loss-aversion	Samuelson and Zeckhauser (1988), Johnson et al. (1993)
Horse race betting	**Favorite-longshot bias**	Favorites are underbet, longshots overbet	Track odds	Single race (not day)	Overweight low p(loss)	Jullien and Salanié (1997)
Horse race betting	**End-of-the-day effect**	Shift to longshots at the end of the day	Track odds	Single day	Reflection effect	McGlothlin (1956)
Insurance	**Buying phone wire insurance**	Consumers buy overpriced insurance	Phone wire insurance purchases	Single wire risk (not portfolio)	Overweight low p(loss)	Cicchetti and Dubin (1994)
Lottery betting	**Demand for Lotto**	More tickets sold as top prize rises	State lottery sales	Single lottery	Overweight low p(win)	Cook and Clotfelter (1993)

1. Finance: The Equity Premium

Two important anomalies in finance can be explained by elements of prospect theory. One anomaly is called the *equity premium*. Stocks—or equities—tend to have more variable annual price changes (or "returns") than bonds do. As a result, the average return to stocks is higher as a way of compensating investors for the additional risk they bear. In most of this century, for example, stock returns were about 8% per year higher than bond returns. This was accepted as a reasonable return premium for equities until Mehra and Prescott (1985) asked how large a degree of risk-aversion is implied by this premium. The answer is surprising: under the standard assumptions of economic theory, investors must be absurdly risk averse to demand such a high premium. For example, a person with enough risk-aversion to explain the equity premium would be indifferent between a coin flip paying either $50,000 or $100,000 and a sure amount of $51,209.

Explaining why the equity premium is so high has preoccupied financial economists for the past 15 years (see Siegel and Thaler 1997). Benartzi and Thaler (1997) suggested a plausible answer based on prospect theory. In their theory, investors are not averse to the variability of returns; they are averse to loss (the chance that returns are negative). Because annual stock returns are negative much more frequently than annual bond returns are, loss-averse investors will demand a large equity premium to compensate them for the much higher chance of losing money in a year. Keep in mind that the higher average return to stocks means that the cumulative return to stocks over a longer horizon is increasingly likely to be positive as the horizon lengthens. Therefore, to explain the equity premium Benartzi and Thaler must assume that investors take a short horizon over which stocks are more likely to lose money than bonds. They compute the expected prospect values of stock and bond returns over various horizons, using estimates of investor utility functions from Kahneman and Tversky (1992) and including a loss-aversion coefficient of 2.25 (i.e., the disutility of a small loss is 2.25 times as large as the utility of an equal gain). Benartzi and Thaler show that over a 1-year horizon, the prospect values of stock and bond returns are about the same if stocks return 8% more than bonds, which explains the equity premium.

Barberis, Huang, and Santos (1999) include loss-aversion in a standard general equilibrium model of asset pricing. They show that loss-aversion and a strong "house money effect" (an increase in risk-preference after stocks have risen) are both necessary to explain the equity premium.

2. Finance: The Disposition Effect

Shefrin and Statman (1985) predicted that because people dislike incurring losses much more than they like incurring gains and are willing to gamble in the domain of losses, investors will hold on to stocks that have lost value (relative to their purchase price) too long and will be eager to sell stocks that have risen in value. They called this the *disposition effect*. The disposition effect is anomalous because the purchase price of a stock should not matter much for whether you decided to sell it. If you think the stock will rise, you should keep it; if you think it will fall, you should sell it. In addition, tax laws encourage people to sell losers rather than winners because such sales generate losses that can be used to reduce the taxes owed on capital gains.

Disposition effects have been found in experiments by Weber and Camerer (1998).[2] On large exchanges, trading volume of stocks that have fallen in price is lower than for stocks that have risen. The best field study was done by Odean (in press). He obtained data from a brokerage firm about all the purchases and sales of a large sample of individual investors. He found that investors held losing stocks a median of 124 days and held winners only 104 days. Investors sometimes say they hold losers because they expect them to "bounce back" (or mean-revert), but in Odean's sample, the unsold losers returned only 5% in the subsequent year, whereas the winners that were sold later returned 11.6%. Interestingly, the winner-loser differences did disappear in December. In this month investors have their last chance to incur a tax advantage from selling losers (and selling winners generates a taxable capital gain), and thus their reluctance to incur losses is temporarily overwhelmed by their last chance to save on taxes.

Genovese and Meyer (in press) report a strong disposition effect in housing sales. Owners who may suffer a nominal loss (selling at a price below what they paid) set prices too high and, as a result, keep their houses too long before selling.

3. Labor Supply

Camerer, Babcock, Loewenstein, and Thaler (in this volume) talked to cab drivers in New York City about when they decide to quit driving each day. Most of the drivers lease their cabs for a fixed fee for up to 12 hours. Many said they set an income target for the day and quit when they reach that target. Although daily income targeting seems sensible, it implies that drivers will work long hours on bad days when the per-hour wage is low and will quit earlier on good high-wage days. The standard theory of the supply of labor predicts the opposite: Drivers will work the hours that are most profitable, quitting early on bad days and making up the shortfall by working longer on good days.

The daily targeting theory and the standard theory of labor supply therefore predict opposite signs of the correlation between hours and the daily wage. To measure the correlation, we collected three samples of data on how many hours drivers worked on different days. The correlation between hours and wages was strongly negative for inexperienced drivers and close to zero for experienced drivers. This suggests that inexperienced drivers began using a daily income

[2] In the Weber and Camerer experiment, subjects whose shares were automatically sold every period (but could be bought back with no transaction cost) did not buy back the shares of losers more than winners. This shows they are not optimistic about the losers but simply reluctant to sell them and lock in a realized loss.

targeting heuristic, but those who did so either tended to quit or learned by experience to shift toward driving around the same number of hours every day.

Daily income targeting assumes loss aversion in an indirect way. To explain why the correlation between hours and wages for inexperienced drivers is so strongly negative, one needs to assume that drivers take a 1-day horizon and have a utility function for the day's income that bends sharply at the daily income target. This bend is an aversion to "losing" by falling short of an income reference point.

4. Asymmetric Price Elasticities of Consumer Goods

The price elasticity of a good is the change in quantity demanded, in percentage terms, divided by the percentage change in its price. Hundreds of studies estimate elasticities by looking at how much purchases change after prices change. Loss-averse consumers dislike price increases more than they like the windfall gain from price cuts and will cut back purchases more when prices rise compared with the extra amount they buy when prices fall. Loss-aversion therefore implies elasticities will be asymmetric, that is, elasticities will be larger in magnitude after price increases than after price decreases. Putler (1992) first looked for such an asymmetry in price elasticities in consumer purchases of eggs and found it.

Hardie, Johnson, and Fader (1993) replicated the study using a typical model of brand choice in which a consumer's utility for a brand is unobserved but can be estimated by observing purchases. They included the possibility that consumers compare a good's current price to a reference price (the last price they paid) and get more disutility from buying when prices have risen than the extra utility they get when prices have fallen. For orange juice, they estimate a coefficient of loss-aversion (the ratio of loss and gain disutilities) around 2.4.

Note that for loss-aversion to explain these results, consumers must be narrowly bracketing purchases of a specific good (e.g., eggs or orange juice). Otherwise, the loss from paying more for one good would be integrated with gains or losses from other goods in their shopping cart and would not loom so large.

5. Savings and Consumption: Insensitivity to Bad Income News

In economic models of lifetime savings and consumption decisions, people are assumed to have separate utilities for consumption in each period, denoted $u[c(t)]$, and discount factors that weight future consumption less than current consumption. These models are used to predict how much rational consumers will consume (or spend) now and how much they will save, depending on their current income, anticipations of future income, and their discount factors. The models make many predictions that seem to be empirically false. The central prediction is that people should plan ahead by anticipating future income to make a guess about their "permanent income" and consume a constant fraction of that total in any one year. Because most workers earn larger and larger incomes throughout their lives, this prediction implies that people will spend more than they earn when they are young—borrowing if they can—and will earn more than they spend when they are older. But in fact, spending on consumption tends to be close to a fixed fraction of current income and does not vary across the life cycle nearly as much as standard theory predicts. Consumption also drops steeply after retirement, which should not be the case if people anticipate retirement and save enough for it.

Shea (1995) pointed out another prediction of the standard life-cycle theory. Think of a group of workers whose wages for the next year are set in advance. In Shea's empirical analysis, these are unionized teachers whose contract is negotiated one-year ahead. In the standard theory, if next year's wage is surprisingly good, then the teachers should spend more now, and if next year's wage is disappointingly low, the teachers should cut back on their spending now. In fact, the teachers in Shea's study did spend more when their future wages were expected to rise, but they *did not* cut back when their future wages were cut.

Bowman, Minehart, and Rabin (1999) can explain this pattern with a stylized two-period consumption-savings model in which workers have reference dependent utility, $u(c(t) - r(t))$ (cf. Duesenberry l949). The utility they get from consumption in each period exhibits loss aversion (the marginal utility of consuming just enough to reach the reference point is always strictly larger than the marginal utility from exceeding it) and a reflection effect (if people are consuming below their reference point, the marginal utility of consumption rises as they get closer to it). Workers begin with some reference point $r(t)$ and save and consume in the first period. Their reference point in the second period is an average of their initial reference point and their first-period consumption, and thus $r(2) = \alpha r(1) + (1 - \alpha)c(1)$. The pleasure workers get from consuming in the second period depends on how much they consumed in the first period through the effect of previous consumption on the current reference point. If they consumed a lot at first, $r(2)$ will be high and they will be disappointed if their standard of living is cut and $c(2) < r(2)$.

Bowman et al. (1999) show formally how this simple model can explain the behavior of the teachers in Shea's study. Suppose teachers are consuming at their reference point and get bad news about future wages (in the sense that the distribution of possible wages next year shifts downward). Bowman et al. show that the teachers may not cut their current consumption at all. Consumption is "sticky downward" for two reasons: (1) Because they are loss-averse, cutting current consumption means they will consume below their reference point this year, which feels awful. (2) Owing to reflection effects, they are willing to gamble that next year's wages might not be so low; thus, they would rather take a gamble in which they either consume far below their reference point or consume right at it than accept consumption that is modestly below the reference point. These two forces make the teachers reluctant to cut their current consumption after receiving bad news about future income prospects, which explains Shea's finding.

6. Status Quo Bias, Endowment Effects, and Buying–Selling Price Gaps

Samuelson and Zeckhauser (1988) coined the term *status quo bias* to refer to an exaggerated preference for the status quo and showed such a bias in a series of experiments. They also reported several observations in field data that are consistent with status quo bias.

When Harvard University added new health-care plan options, older faculty members who were hired previously when the new options were not available were, of course, allowed to switch to the new options. If one assumes that the new and old faculty members have essentially the same preferences for health-care plans, then the distribution of plans elected by new and old faculty should be the same. However, Samuelson and Zeckhauser found that older faculty members tended to stick to their previous plans; compared with the newer faculty members, fewer of the old faculty elected new options.

In cases in which there is no status quo, people may have an exaggerated preference for whichever option is the default choice. Johnson, Hershey, Meszaros, and Kunreuther (1993) observed this phenomenon in decisions involving insurance purchases. At the time of their study, Pennsylvania and New Jersey legislators were considering various kinds of tort reform allowing firms to offer cheaper automobile insurance that limited the rights of the insured person to sue for damages from accidents. Both states adopted very similar forms of limited insurance, but they chose different default options, creating a natural experiment. All insurance companies mailed forms to their customers asking them whether they wanted the cheaper limited-rights insurance or the more expensive unlimited-rights insurance. One state made the limited-rights insurance the default (the insured person would get that if they did not respond), and the other made unlimited-rights the default. In fact, the percentage of people actively electing the limited-rights insurance was higher in the state where that was the default. An experiment replicated the effect.

A closely related body of research on endowment effects established that buying and selling prices for a good are often quite different. The paradigmatic experimental demonstration of this is the "mugs" experiments of Kahneman, Knetsch, and Thaler (1990). In their experiments, some subjects are endowed (randomly) with coffee mugs, and others are not. Those who are given the mugs demand a price about 2–3 times as large as the price that those without mugs are willing to pay, even though in economic theory these prices should be extremely close together. In fact, the mug experiments were inspired by field observations of large gaps in hypothetical buying and selling prices in "contingent valuations." Contingent valuations are measurements of the economic value of goods that are not normally traded—like clean air, environmental damage, and so forth. These money valuations are used for doing benefit-cost analysis and establishing economic damages in lawsuits. There is a huge literature establishing that selling prices are generally much larger than buying prices, although there is a heated debate among psychologists and economists about what the price gap means and how to measure "true" valuations in the face of such a gap.

All three phenomena (status quo biases default preference, and endowment effects) are consistent with aversion to losses relative to a reference point. Making one option the status quo or default or endowing a person with a good (even hypothetically) seems to establish a reference point people move away from only reluctantly, or if they are paid a large sum.

7. Racetrack Betting: The Favorite-Longshot Bias

In parimutuel betting on horse races, there is a pronounced bias toward betting on "longshots," which are horses with a relatively small chance of winning. That is, if one groups longshots with the same percentage of money bet on them into a class, the fraction of time horses in that class win is far smaller than the percentage of money bet on them. Longshot horses with 2% of the total money bet on them, for example, win only about 1% of the time (see Thaler and Ziemba 1988; Hausch and Ziemba 1995).

Overbetting longshots implies favorites are underbet. Indeed, some horses are so heavily favored that up to 70% of the win money is wagered on them. For these heavy favorites, the return for a dollar bet is very low if the horse wins. (Because the track keeps about 15% of the money bet for expenses and profit, bettors who bet on such a heavy favorite share only 85% of the money with 70% of the people, which results in a payoff of only about $2.40 for a $2 bet.) People dislike these bets so much that, in fact, if one makes those bets it is possible to earn a small positive profit (even accounting for the track's 15% take).

There are many explanations for the favorite-longshot bias, each of which probably contributes to the phenomenon. Horses that have lost many races in a row tend to be longshots, and thus a gambler's fallacious belief that such horses are due for a win may contribute to overbetting on them. Prospect-theoretic overweighting of low probabilities of winning will also lead to overbetting of longshots.

Within standard expected utility theory, the favorite-longshot bias can only be explained by assuming that people have convex utility functions for money outcomes. The most careful study comparing expected utility and prospect theory was done by Jullien and Salanié (1997). Their study used a huge sample of all the flat races run in England for ten years (34,443 races). They assumed that bettors value bets on horses by using either expected-utility theory, rank-dependent utility theory, or cumulative prospect theory (see Kahneman and Tversky 1992). If the marginal bettor is indifferent among bets on all the horses at the odds established when the race is run, then indifference conditions can be used to infer the parameters of that bettor's utility and probability weighting functions.

Jullien and Salanié found that cumulative prospect theory fits much better than rank-dependent theory and expected utility theory. They estimated that the utility function for small money amounts is convex. Their estimate of the probability weighting function $\pi(p)$ for probabilities of gain is almost linear, but the weighting function for loss probabilities severely overweights low probabilities of loss (e.g., $\pi(.1) = .45$ and $\pi(.3) = .65$). These estimates imply a surprising new explanation for the favorite-longshot bias: Bettors like longshots because they have convex utility and weight their high chances of losing and small chances of winning roughly linearly. They hate favorites, however, because they like to gamble ($u(x)$ is convex) but are disproportionately afraid of the small chance of losing when they bet on a heavy favorite. (In my personal experience as a betting researcher, I have found that losing on a heavy favorite is particularly disappointing—an emotional effect the Jullien-Salanié estimates capture.)

8. Racetrack Betting: The End-of-the-Day Effect

McGlothlin (1956) and Ali (1977) established another racetrack anomaly that points to the central role of reference points. They found that bettors tend to shift their bets toward longshots, and away from favorites, later in the racing day. Because the track takes a hefty bite out of each dollar, most bettors are behind by the last race of the day. These bettors really prefer longshots because a small longshot bet can generate a large enough profit to cover their earlier losses, enabling them to break even. The movement toward longshots, and away from favorites, is so pronounced that some studies show that conservatively betting on the favorite to show (to finish first, second, or third) in the last race is a profitable bet despite the track's take.

The end-of-the-day effect is consistent with using zero daily profit as a reference point and gambling in the domain of losses to break even. Expected-utility theory cannot gracefully explain the shift in risk preferences across the day if bettors integrate their wealth because the last race on a Saturday is not fundamentally different than the first race on the bettor's next outing. Cumulative prospect theory can explain the shift by assuming people open a mental account at the beginning of the day, close it at the end, and hate closing an account in the red.

9. Telephone Wire Repair Insurance

Ciccheti and Dubin (1994) conducted an interesting study of whether people purchase insurance against damage to their telephone wiring. The phone companies they studied either required customers to pay for the cost of wiring repair, about \$60, or to buy insurance for \$.45 per month. Given phone company estimates of the frequency of wire damage, the expected cost of wire damage is only \$.26.

Ciccheti and Dubin looked across geographical areas with different probabilities of wire damage rates to see whether cross-area variation in the tendency to buy insurance was related to different probabilities. They did find a relation and exploited this to estimate parameters of an expected-utility model. They found some evidence that people were weighting damage probabilities nonlinearly and also some evidence of status quo bias. (People who had previously been uninsured, when a new insurance option was introduced, were less likely to buy it than new customers were.)

More importantly, Ciccheti and Dubin never asked whether it is reasonable to purchase insurance against such a tiny risk. In standard expected utility, a person who is averse to very modest risks at all levels of wealth should be more risk-averse to large risks. Rabin (in press) was the first to demonstrate how dramatic the implications of local risk-aversion are for global risk-aversion. He showed formally that a mildly risk-averse expected-utility maximizer who would turn down a coin flip (at all wealth levels) in which he or she is equally likely to win \$11 or lose \$10 should not accept a coin flip in which \$100, could be lost, *regardless of how much he or she could win*. In expected utility terms, turning down the small-stakes flip implies a little bit of curvature in a \$21 range of a concave utility function. Turning down the small-stakes flip for all wealth levels implies the utility function is slightly curved at all wealth levels, which mathematically implies a dramatic degree of global curvature.

Rabin's proof suggests a rejection of the joint hypotheses that consumers who buy wire repair insurance are integrating their wealth and valuing the insurance according to expected utility (and know the correct probabilities of damage). A more plausible explanation comes immediately from prospect theory—consumers are overweighting the probability of damage. (Loss-aversion and reflection cannot explain their purchases because, if they are loss averse, they should dislike spending the \$.45 per month, and reflection implies they will never insure unless they overestimate the probability of loss.) Once again, narrow bracketing is also required: consumers must be focusing only on wire repair risk; otherwise, the tiny probability of a modest loss would be absorbed into a portfolio of life's ups and downs and weighted more reasonably.

10. State Lotteries

Lotto is a special kind of lottery game in which players choose six different numbers from a set of 40–50 numbers. They win a large jackpot if their six choices match six numbers that are randomly drawn in public. If no player picks all six numbers correctly, the jackpot is rolled over and added to the next week's jackpot; several weeks of rollovers can build up jackpots up to \$350 million or more. The large jackpots have made lotto very popular.[3] Lotto was introduced in several American states in 1980 and accounted for about half of all state lottery ticket sales by 1989.

[3] A similar bet, the "pick six," was introduced at horse-racing tracks in the 1980s. In the pick six, bettors must choose the winners of six races. This is extremely hard to do, and thus a large rollover occurs if nobody has picked all six winners several days in a row, just like lotto. Pick-six betting now accounts for a large fraction of overall betting.

Cook and Clotfelter (1993) suggest that the popularity of Lotto results from players' being more sensitive to the large jackpot than to the correspondingly probability of winning. They write,

> If players tend to judge the likelihood of winning based on the frequency with which someone wins, then a larger state can offer a game at longer odds but with the same perceived probability of winning as a smaller state. The larger population base in effect conceals the smaller probability of winning the jackpot, while the larger jackpot is highly visible. This interpretation is congruent with prospect theory. (p. 634)

Their regressions show that across states, ticket sales are strongly correlated with the size of a state's population (which is correlated with jackpot size). Within a state, ticket sales each week are strongly correlated with the size of the rollover. In expected utility, this can be explained only by utility functions for money that are convex. Prospect theory easily explains the demand for high jackpots, as Cook and Clotfelter suggest, by overweighting of, and insensitivity toward, very low probabilities.

Conclusions

Economists value (1) mathematical formalism and econometric parsimony, and (2) the ability of theory to explain naturally occurring data. I share these tastes. This article has demonstrated that prospect theory is valuable in both ways because it can explain ten patterns observed in a wide variety of economic domains with a small number of modeling features. Different features of prospect theory help explain different patterns. *Loss-aversion* can explain the extra return on stocks compared with bonds (the equity premium), the tendency of cab drivers to work longer hours on low-wage days, asymmetries in consumer reactions to price increases and decreases, the insensitivity of consumption to bad news about income, and status quo and endowment effects. *Reflection effects*—gambling in the domain of a perceived loss—can explain holding losing stocks longer than winners and refusing to sell your house at a loss (disposition effects), insensitivity of consumption to bad income news, and the shift toward longshot betting at the end of a racetrack day. *Nonlinear weighting of probabilities* can explain the favorite-longshot bias in horse-race betting, the popularity of lotto lotteries with large jackpots, and the purchase of telephone wire repair insurance. In addition, note that the disposition effect and downward-sloping labor supply of cab drivers were not simply observed but were also predicted in advance based on prospect theory.

In all these examples it is also necessary to assume people are isolating or narrowly bracketing the relevant decisions. Bracketing narrowly focuses attention most dramatically on the possibility of a loss or extreme outcome, or a low probability. With broader bracketing, outcomes are mingled with other gains and losses, diluting the psychological influence of any single outcome and making these phenomena hard to explain as a result of prospect theory valuation.

I have two final comments. First, I have chosen examples in which there are several studies, or one very conclusive one, showing regularities in field data that cannot be easily reconciled with expected utility theory. However, these regularities can be explained by adding extra assumptions. The problem is that these extras are truly ad hoc because each regularity requires a special assumption. Worse, an extra assumption that helps explain one regularity may contradict another. For example, assuming people are risk-preferring (or have convex utility for money) can explain the popularity of longshot horses and lotto, but that assumption predicts stocks should return *less* than bonds, which is wildly false. You can explain why cab drivers drive long hours on bad days by assuming they cannot borrow (they are liquidity constrained), but liquidity constraint implies teachers who get good income news should not be able to spend more, whereas those who get bad news can cut back, which is exactly the opposite of what they do.

Second, prospect theory is a suitable replacement for expected utility because it can explain anomalies like those listed above and can *also* explain the most basic phenomena expected utility is used to explain. A prominent example is pricing of financial assets discussed above in sections 1 and 2. Another prominent example, which appears in every economics textbook, is the voluntary purchase of insurance by people. The expected utility explanation for why people buy actuarially unfair insurance is that they have concave utility, and thus they hate losing large amounts of money disproportionally compared with spending small amounts on insurance premiums.

In fact, many people *do not* purchase insurance voluntarily (e.g., most states require automobile insurance by law). The failure to purchase is inconsistent with the expected utility explanation and more easy to reconcile with prospect theory (because the disutility of loss is assumed to be convex). When people *do* buy insurance, people are probably avoiding low-probability disasters that they overweight (the prospect theory explanation) rather than avoiding a steep drop in a concave utility function (the expected utility theory explanation).

A crucial kind of evidence that distinguishes the two explanations comes from experiments on probabilistic insurance, which is insurance that does *not* pay a claim, if an accident occurs, with some probability r. According to expected utility theory, if r is small, people should pay approximately $(1 - r)$ times as much for probabilistic insurance as they pay for full insurance (Wakker, Thaler, and Tversky 1997). But experimental responses show that people hate probabilistic insurance; they pay a multiple much less than $1 - r$ for it (for example, they pay 80% as much when $r = .01$ when they should pay 99% as much). Prospect theory can explain their hatred easily: probabilistic insurance does not reduce the probability of loss all the way toward zero, and the low probability r is still overweighted. Prospect theory can therefore explain why people buy full insurance *and* why they do not buy probabilistic insurance. Expected utility cannot do both.

Because prospect theory can explain the basic phenomena expected utility was most fruitfully applied to, like asset pricing and insurance purchase, and can also explain field anomalies like the ten listed in table 5.1 (two of which were predicted), there is no good scientific reason why it should not be used alongside expected utility in current research and be given prominent space in economics textbooks.

References

Ali, Muhktar. 1977. "Probability and Utility Estimates for Racetrack Bettors." *Journal of Political Economy*, 85: 803–15.

Benartzi, Shlomo, and Richard Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics*, 110: 73–92.

Bowman, David, Debby Minehart, and Matthew Rabin. 1997. "Loss Aversion in a Savings Model." Working paper, University of California, Berkeley.

Camerer, Colin F. 1995. "Individual Decision Making." In *Handbook of Experimental Economics*, edited by A. E. Roth and J. Kagel. Princeton: Princeton University Press.

Ciccheti, Charles, and Jeff Dubin. 1994. "A Microeconometric Analysis of Risk-Aversion and the Decision to Self-insure." *Journal of Political Economy*, 102: 169–86.

Cook, Philip I., and Charles T. Clotfelter. 1993. "The Peculiar Scale Economies of Lotto." *American Economic Review*, 83 (June): 634–43.

Hardie, Bruce G. S., Eric J. Johnson, and Peter S. Fader. 1993. "Modeling Loss Aversion and Reference Dependence Effects on Brand Choice." *Marketing Science*.

Hausch, Donald B., and William T. Ziemba. 1995. "Efficiency in Sports and Lottery Betting Markets." In *Handbook of Finance*, edited by R. A. Jarrow, V. Maksimovic, and W. T. Ziemba. Amsterdam: North-Holland.

Johnson, Eric, Jack Hershey, Jacqueline Meszaros, and Howard Kunreuther. 1993. "Framing, Probability Distortions, and Insurance Decisions." *Journal of Risk and Uncertainty*, 7: 35–51.

Jullien, Bruno, and Bernard Salanié. 1997. "Estimating Preferences under Risk: The Case of Racetrack Bettors. IDEI and GREMAQ Toulouse University working paper.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*, 98: 1325–48.

Kahneman, Daniel, and Amos Tversky. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5 (October): 297–324.

McGlothlin, William H. 1956. "Stability of Choices among Uncertain Alternatives." *American Journal of Psychology*, 69: 604–15.

Mehra, Rajneesh, and Edward Prescott. 1985. "The Equity Premium Puzzle." *Journal of Monetary Economics*, 15: 145–61.

Odean, Terrance. 1998. "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance* 53(5): 1775–98.

Putler, Daniel S. 1992. "Incorporating Reference Price Effects into a Theory of Consumer Choice." *Marketing Science*, 11(3): 287–309.

Rabin, Matthew. 1998. "Psychology and Economics." *Journal of Economic Literature*, 36: 11–46.

Rabin, Matthew. 2000. "Risk-aversion and expected utility: A calibration theorem." *Econometrica*, 68: 1281–92.

Read, Daniel, George Loewenstein, and Matthew Rabin. 1998. "Choice bracketing." Work-

ing paper, Department of Social and Decision Sciences, Carnegie-Mellon University.
Samuelson, William, and Richard Zeckhauser. 1988. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty*, 1: 7–59.
Shea, John. 1995. "Union Contracts and the Life-Cycle/Permanent-Income Hypothesis." *American Economic Review*, 85: 186–200.
Shefrin, Hersh, and Meier Statman. 1985. "The Disposition to Sell Winners too Early and Ride Losers too Long." *Journal of Finance*, 40: 777–90.
Siegel, Jeremy, and Richard Thaler. 1997. "The Equity Premium Puzzle." *Journal of Economic Perspectives*, 11(Winter): 191–200.
Thaler, Richard, and William T. Ziemba. 1988. "Parimutuel Betting Markets: Racetracks and Lotteries." *Journal of Economic Perspectives*, 2: 161–74.
Wakker, Peter P., Richard H. Thaler, and Amos Tversky. 1997. "Probabilistic Insurance." *Journal of Risk and Uncertainty*, 15: 5–26.
Weber, Martin, and Colin Camerer. In press. "The Disposition Effect in Securities Trading." *Journal of Economic Behavior and Organization*.

C H A P T E R 6

Time Discounting and Time Preference: A Critical Review

SHANE FREDERICK, GEORGE LOEWENSTEIN, AND TED O'DONOGHUE

INTERTEMPORAL choices—decisions involving trade-offs among costs and benefits occurring at different times—are important and ubiquitous. Such decisions not only affect one's health, wealth, and happiness, but may also, as Adam Smith first recognized, determine the economic prosperity of nations. In this chapter, we review empirical research on intertemporal choice, and present an overview of recent theoretical formulations that incorporate insights gained from this research.

Economists' attention to intertemporal choice began early in the history of the discipline. Not long after Adam Smith called attention to the importance of intertemporal choice for the wealth of nations, the Scottish economist John Rae was examining the sociological and psychological determinants of these choices. We will briefly review the perspectives on intertemporal choice of Rae and nineteenth- and early twentieth-century economists, and describe how these early perspectives interpreted intertemporal choice as the joint product of many conflicting psychological motives.

All of this changed when Paul Samuelson proposed the discounted-utility (DU) model in 1937. Despite Samuelson's manifest reservations about the normative and descriptive validity of the formulation he had proposed, the DU model was accepted almost instantly, not only as a valid normative standard for public policies (for example, in cost-benefit analyses), but as a descriptively accurate representation of actual behavior. A central assumption of the DU model is that all of the disparate motives underlying intertemporal choice can be condensed into a single parameter—the discount rate. We do not present an axiomatic derivation of the DU model, but instead focus on those features that highlight the implicit psychological assumptions underlying the model.

We thank John McMillan, David Laibson, Colin Camerer, Nachum Sicherman, Duncan Simester, and three anonymous referees for useful comments. We thank Cara Barber, Rosa Blackwood, Mandar Oak, and Rosa Stipanovic for research assistance. For financial support, Frederick and Loewenstein thank the Integrated Study of the Human Dimensions of Global Change at Carnegie Mellon University (NSF Grant SBR-9521914), and O'Donoghue thanks the National Science Foundation (Award SES-0078796). This chapter was reprinted with the permission of the American Economic Association. It was originally published in 2002 as "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature* 40(June): 351–401.

Samuelson's reservations about the descriptive validity of the DU model were justified. Virtually every assumption underlying the DU model has been tested and found to be descriptively invalid in at least some situations. Moreover, these anomalies are not anomalies in the sense that they are regarded as errors by the people who commit them. Unlike many of the better-known expected-utility anomalies, the DU anomalies do not necessarily violate any standard or principle that people believe they should uphold.

The insights about intertemporal choice gleaned from this empirical research have led to the proposal of numerous alternative theoretical models. Some of these modify the discount function, permitting, for example, declining discount rates or "hyperbolic discounting." Others introduce additional arguments into the utility function, such as the utility of anticipation. Still others depart from the DU model more radically, by including, for instance, systematic mispredictions of future utility. Many of these new theories revive psychological considerations discussed by Rae and other early economists that were extinguished with the adoption of the DU model and its expression of intertemporal preferences in terms of a single parameter.

While the DU model assumes that people are characterized by a single discount rate, the literature reveals spectacular variation across (and even within) studies. The failure of this research to converge toward any agreed-upon average discount rate stems partly from differences in elicitation procedures. But it also stems from the faulty assumption that the varied considerations that are relevant in intertemporal choices apply equally to different choices and thus that they can all be sensibly represented by a single discount rate.

Throughout, we stress the importance of distinguishing among the varied considerations that underlie intertemporal choices. We distinguish *time discounting* from *time preference*. We use the term *time discounting* broadly to encompass *any* reason for caring less about a future consequence, including factors that diminish the expected utility generated by a future consequence, such as uncertainty or changing tastes. We use the term *time preference* to refer, more specifically, to the preference for immediate utility over delayed utility. We push this theme further by examining whether time preference itself might consist of distinct psychological traits that can be separately analyzed.

Historical Origins of the Discounted-Utility Model

The historical developments that culminated in the formulation of the DU model help to explain the model's limitations. Each of the major figures in its development—John Rae, Eugen von Böhm-Bawerk, Irving Fisher, and Paul Samuelson—built upon the theoretical framework of his predecessors, drawing on little more than introspection and personal observation. When the DU model eventually became entrenched as the dominant theoretical framework for modeling intertemporal choice, it was due largely to its simplicity and its resemblance to the familiar compound interest formula, and not as a result of empirical research demonstrating its validity.

Intertemporal choice became firmly established as a distinct topic in 1834, with John Rae's publication of *The Sociological Theory of Capital*. Like Adam Smith, Rae sought to determine why wealth differed among nations. Smith had argued that national wealth was determined by the amount of labor allocated to the production of capital, but Rae recognized that this account was incomplete because it failed to explain the determinants of this allocation. In Rae's view, the missing element was "the effective desire of accumulation"—a psychological factor that differed across countries and determined a society's level of saving and investment.

Along with inventing the topic of intertemporal choice, Rae also produced the first in-depth discussion of the psychological motives underlying intertemporal choice. Rae believed that intertemporal choices were the joint product of factors that either promoted or limited the effective desire of accumulation. The two main factors that promoted the effective desire of accumulation were the bequest motive—"the prevalence throughout the society of the social and benevolent affections"—and the propensity to exercise self-restraint: "the extent of the intellectual powers, and the consequent prevalence of habits of reflection, and prudence, in the minds of the members of society" (Rae 1905 [1834], p. 58). One limiting factor was the uncertainty of human life:

> When engaged in safe occupations, and living in healthy countries, men are much more apt to be frugal, than in unhealthy, or hazardous occupations, and in climates pernicious to human life. Sailors and soldiers are prodigals. In the West Indies, New Orleans, the East Indies, the expenditure of the inhabitants is profuse. The same people, coming to reside in the healthy parts of Europe, and not getting into the vortex of extravagant fashion, live economically. War and pestilence always have waste and luxury, among the other evils that follow in their train (Rae 1905 [1834], p. 57).

A second factor that limited the effective desire of accumulation was the excitement produced by the prospect of immediate consumption, and the concomitant discomfort of deferring such available gratifications:

> Such pleasures as may now be enjoyed generally awaken a passion strongly prompting to the partaking of them. The actual presence of the immediate object of desire in the mind by exciting the attention, seems to rouse all the faculties, as it were, to fix their view on it, and leads them to a very lively conception of the enjoyments which it offers to their instant possession (Rae 1905 [1834], p. 120).

Among the four factors that Rae identified as the joint determinants of time preference, one can glimpse two fundamentally different views. One, which was later championed by William S. Jevons (1888) and his son, Herbert S. Jevons (1905), assumes that people care only about their immediate utility, and explains farsighted behavior by postulating utility from the anticipation of future consumption. In this view, deferral of gratification will occur only if it produces an increase in "anticipal" utility that more than compensates for the decrease in immediate consumption utility. The second perspective assumes equal treatment of present and future (zero discounting) as the natural baseline for behavior, and attributes the overweighting of the present to the miseries produced by the self-denial required to delay gratification. N. W. Senior, the best-known advocate of this "abstinence" perspective, wrote, "To abstain from the enjoyment which is in our power, or to seek distant rather than immediate results, are among the most painful exertions of the human will" (Senior 1836, 60).

The anticipatory-utility and abstinence perspectives share the idea that intertemporal trade-offs depend on immediate feelings—in one case, the immediate pleasure of anticipation, and in the other, the immediate discomfort of self-denial. The two perspectives, however, explain variability in intertemporal-choice behavior in different ways. The anticipatory-utility perspective attributes variations in intertemporal-choice behavior to differences in people's abilities to imagine the future and to differences in situations that promote or inhibit such mental images. The abstinence perspective, on the other hand, explains variations in intertemporal-choice behavior on the basis of individual and situational differences in the psychological discomfort associated with self-denial. In this view, one should observe high rates of time discounting by people who find it painful to delay gratification, and in situations in which deferral is generally painful—for example, when one is, as Rae worded it, in the "actual presence of the immediate object of desire."

Eugen von Böhm-Bawerk, the next major figure in the development of the economic perspective on intertemporal choice, added a new motive to the list proposed by Rae, Jevons, and Senior, arguing that humans suffer from a systematic tendency to underestimate future wants.

> It may be that we possess inadequate power to imagine and to abstract, or that we are not willing to put forth the necessary effort, but in any event we limn a more or less incomplete picture of our future wants and especially of the remotely distant ones. And then there are all those wants that never come to mind at all.[1] (Böhm-Bawerk 1970 [1889], 268–69)

Böhm-Bawerk's analysis of time preference, like those of his predecessors, was heavily psychological, and much of his voluminous treatise, *Capital and Interest*, was devoted to discussions of the psychological constituents of time preference. However, whereas the early views of Rae, Senior, and Jevons explained intertemporal choices in terms of motives uniquely associated with time, Böhm-Bawerk began modeling intertemporal choice in the same terms as other economic trade-offs—as a "technical" decision about allocating resources (to oneself) over different points in time, much as one would allocate resources between any two competing interests, such as housing and food.

Böhm-Bawerk's treatment of intertemporal choice as an allocation of consumption among time periods was formalized a decade later by the American

[1] In a frequently cited passage from *The Economics of Welfare*, Arthur Pigou (1920, p. 25) proposed a similar account of time preference, suggesting that it results from a type of cognitive illusion: "our telescopic faculty is defective, and we, therefore, see future pleasures, as it were, on a diminished scale."

economist Irving Fisher (1930). Fisher plotted the intertemporal consumption decision on a two-good indifference diagram, with consumption in the current year on the abscissa, and consumption in the following year on the ordinate. This representation made clear that a person's observed (marginal) rate of time preference—the marginal rate of substitution at her chosen consumption bundle—depends on two considerations: time preference and diminishing marginal utility. Many economists have subsequently expressed discomfort with using the term *time preference* to include the effects of differential marginal utility arising from unequal consumption levels between time periods (see in particular Olson and Bailey 1981). In Fisher's formulation, *pure* time preference can be interpreted as the marginal rate of substitution on the diagonal, where consumption is equal in both periods.

Fisher's writings, like those of his predecessors, included extensive discussions of the psychological determinants of time preference. Like Böhm-Bawerk, he differentiated "objective factors," such as projected future wealth and risk, from "personal factors." Fisher's list of personal factors included the four described by Rae, "foresight" (the ability to imagine future wants—the inverse of the deficit that Böhm-Bawerk postulated), and "fashion," which Fisher believed to be "of vast importance . . . in its influence both on the rate of interest and on the distribution of wealth itself" (Fisher 1930, p. 88). He wrote,

> The most fitful of the causes at work is probably fashion. This at the present time acts, on the one hand, to stimulate men to save and become millionaires, and, on the other hand, to stimulate millionaires to live in an ostentatious manner. (p. 87)

Hence, in the early part of the twentieth century, "time preference" was viewed as an amalgamation of various intertemporal motives. While the DU model condenses these motives into the discount rate, we will argue that resurrecting these distinct motives is crucial for understanding intertemporal choices.

The Discounted-Utility Model

In 1937, Paul Samuelson introduced the DU model in a five-page article titled "A Note on Measurement of Utility." Samuelson's paper was intended to offer a generalized model of intertemporal choice that was applicable to multiple time periods (Fisher's graphical indifference-curve analysis was difficult to extend to more than two time periods) and to make the point that representing intertemporal trade-offs required a cardinal measure of utility. But in Samuelson's simplified model, all the psychological concerns discussed in the previous century were compressed into a single parameter, the discount rate.

The DU model specifies a decision maker's intertemporal preferences over consumption profiles ($c_t, \ldots, c_T$). Under the usual assumptions (completeness, transitivity, and continuity), such preferences can be represented by an intertemporal utility function $U^t(c_t, \ldots, c_T)$. The DU model goes further, by assuming that a person's intertemporal utility function can be described by the following special functional form:

$$U^t(c_t,\ldots,c_T)=\sum_{k=0}^{T-t}D(k)u(c_{t+k})\quad \text{where } D(k)=\left(\frac{1}{1+\rho}\right)^k.$$

In this formulation, $u(c_{t+k})$ is often interpreted as the person's cardinal instantaneous utility function—her well-being in period $t+k$—and $D(k)$ is often interpreted as the person's discount function—the relative weight that she attaches, in period t, to her well-being in period $t+k$. ρ represents the individual's pure rate of time preference (her discount rate), which is meant to reflect the collective effects of the "psychological" motives discussed earlier.[2]

Samuelson did not endorse the DU model as a normative model of intertemporal choice, noting that "any connection between utility as discussed here and any welfare concept is disavowed" (1937, p. 161). He also made no claims on behalf of its descriptive validity, stressing, "It is completely arbitrary to assume that the individual behaves so as to maximize an integral of the form envisaged in [the DU model]" (p. 159). Yet despite Samuelson's manifest reservations, the simplicity and elegance of this formulation was irresistible, and the DU model was rapidly adopted as the framework of choice for analyzing intertemporal decisions.

The DU model received a scarcely needed further boost to its dominance as the standard model of intertemporal choice when Tjalling C. Koopmans (1960) showed that the model could be derived from a superficially plausible set of axioms. Koopmans, like Samuelson, did not argue that the DU model was psychologically or normatively plausible; his goal was only to show that under some well-specified (though arguably unrealistic) circumstances, individuals were logically compelled to possess positive time preference. Producers of a product, however, cannot dictate how the product will be used, and Koopmans's central technical message was largely lost while his axiomatization of the DU model helped to cement its popularity and bolster its perceived legitimacy.

We next describe some important features of the DU model as it is commonly used by economists, and briefly comment on the normative and positive validity of these assumptions. These features do not represent an axiom system—they are neither necessary nor sufficient conditions for the DU model—but are intended to highlight the implicit psychological assumptions underlying the model.[3]

Integration of New Alternatives with Existing Plans

A central assumption in most models of intertemporal choice—including the DU model—is that a person evaluates new alternatives by integrating them with one's

[2] The continuous-time analogue is $U^t(\{c_\tau\}_{\tau\in[t,T]})=\int_{\tau=t}^{T}e^{-\rho(\tau-1)}u(c_\tau)$. For expositional ease, we shall restrict attention to discrete-time throughout.

[3] There are several different axiom systems for the DU model—in addition to Koopmans, see Fishburn (1970), Lancaster (1963), Meyer (1976), and Fishburn and Rubinstein (1982).

existing plans. To illustrate, consider a person with an existing consumption plan $(c_t, \ldots, c_T)$ who is offered an intertemporal-choice prospect X, which might be something like an option to give up \$5,000 today to receive \$10,000 in five years. Integration means that prospect X is not evaluated in isolation, but in light of how it changes the person's aggregate consumption in all future periods. Thus, to evaluate the prospect X, the person must choose what his or her new consumption path $(c'_t, \ldots, c'_T)$ would be if he or she were to accept prospect X, and should accept the prospect if $U^t(c'_t, \ldots, c'_T) > U^t(c_t, \ldots, c_T)$.

An alternative way to understand integration is to recognize that intertemporal prospects alter a person's budget set. If the person's initial endowment is E_0, then accepting prospect X would change his or her endowment to $E_0 \cup X$. Letting $B(E)$ denote the person's budget set given endowment E—that is, the set of consumption streams that are feasible given endowment E—the DU model says that the person should accept prospect X if:

$$\max_{(c_t,\ldots,c_T)\in B(E_0\cup X)} \sum_{\tau=t}^{T}\left(\frac{1}{1+\rho}\right)^{\tau-t} u(c_\tau) > \max_{(c_t,\ldots,c_T)\in B(E_0)} \sum_{\tau=t}^{T}\left(\frac{1}{1+\rho}\right)^{\tau-t} u(c_\tau).$$

While integration seems normatively compelling, it may be too difficult actually to do. A person may not have well-formed plans about future consumption streams, or be unable (or unwilling) to recompute the new optimal plan every time he or she makes an intertemporal choice. Some of the evidence we will review supports the plausible presumption that people evaluate the results of intertemporal choices independently of any expectations they have regarding consumption in future time periods.

Utility Independence

The DU model explicitly assumes that the overall value—or "global utility"—of a sequence of outcomes is equal to the (discounted) sum of the utilities in each period. Hence, the distribution of *utility* across time makes no difference beyond that dictated by discounting, which (assuming positive time preference) penalizes utility that is experienced later. The assumption of utility independence has rarely been discussed or challenged, but its implications are far from innocuous. It rules out any kind of preference for patterns of utility over time—for example, a preference for a flat utility profile over a roller-coaster utility profile with the same discounted utility.[4]

[4] "Utility independence" has meaning only if one literally interprets $u(c_{t+k})$ as well-being experienced in period $t + k$. We believe that this is, in fact, the common interpretation. For a model that relaxes the assumption of utility independence see Hermalin and Isen (2000), who consider a model in which well-being in period t depends on well-being in period $t - 1$—that is, they assume $u_t = u(c_t, u_{t-1})$. See also Kahneman, Wakker, and Sarin (1997), who propose a set of axioms that would justify an assumption of additive separability in instantaneous utility.

Consumption Independence

The DU model explicitly assumes that a person's well-being in period $t + k$ is independent of his or her consumption in any other period—that is, that the marginal rate of substitution between consumption in periods τ and τ' is independent of consumption in period τ''.

Consumption independence is analogous to, but fundamentally different from, the independence axiom of expected-utility theory. In expected-utility theory, the independence axiom specifies that preferences over uncertain prospects are not affected by the consequences that the prospects share—that is, that the utility of an experienced outcome is unaffected by other outcomes that one might have experienced (but did not). In intertemporal choice, consumption independence says that preferences over consumption profiles are not affected by the nature of consumption in periods in which consumption is identical in the two profiles—that is, that an outcome's utility is unaffected by outcomes experienced in prior or future periods. For example, consumption independence says that one's preference between an Italian and Thai restaurant tonight should not depend on whether one had Italian last night nor whether one expects to have it tomorrow. As the example suggests, and as Samuelson and Koopmans both recognized, there is no compelling rationale for such an assumption. Samuelson (1952, p. 674) noted that "the amount of wine I drank yesterday and will drink tomorrow can be expected to have effects upon my today's indifference slope between wine and milk." Similarly, Koopmans (1960, p. 292) acknowledged, "One cannot claim a high degree of realism for [the independence assumption], because there is no clear reason why complementarity of goods could not extend over more than one time period."

Stationary Instantaneous Utility

When applying the DU model to specific problems, it is often assumed that the cardinal instantaneous utility function $u(c_\tau)$ is constant across time, so that the well-being generated by any activity is the same in different periods. Most economists would acknowledge that stationarity of the instantaneous utility function is not sensible in many situations, because people's preferences in fact do change over time in predictable and unpredictable ways. Though this unrealistic assumption is often retained for analytical convenience, it becomes less defensible as economists gain insight into how tastes change over time (see Loewenstein and Angner, in press, for a discussion of different sources of preference change).[5]

[5] As will be discussed, endogenous preference changes, due to things such as habit formation or reference dependence, are best understood in terms of consumption interdependence and not nonstationary utility. In some situations, nonstationarities clearly play an important role in behavior—see, for example, Suranovic, Goldfarb, and Leonard (1999) and O'Donoghue and Rabin (1999a, 2000) discuss the importance of nonstationarities in the realm of addictive behavior.

Independence of Discounting from Consumption

The DU model assumes that the discount function is invariant across all forms of consumption. This feature is crucial to the notion of *time* preference. If people discount utility from different sources at different rates, then the notion of a unitary time preference is meaningless. Instead we would need to label time preference according to the object being delayed—“banana time preference,” “vacation time preference,” and so on.

Constant Discounting and Time Consistency

Any discount function can be written in the form

$$D(k) = \prod_{n=0}^{k-1}\left(\frac{1}{1+\rho_n}\right),$$

where ρ_n represents the per-period discount rate for period n—that is, the discount rate applied between periods n and $n + 1$. Hence, by assuming that the discount function takes the form

$$D(k) = \left(\frac{1}{1+\rho}\right)^k,$$

the DU model assumes a constant per-period discount rate ($\rho_n = \rho$ for all n).[6]

Constant discounting entails an evenhandedness in the way a person evaluates time. It means that delaying or accelerating two dated outcomes by a common amount should not change preferences between the outcomes—if in period t one prefers X at τ to Y at $\tau + d$ for *some* τ, then in period t one must prefer X at τ to Y at $\tau + d$ for *all* τ. The assumption of constant discounting permits a person's time preference to be summarized as a single discount *rate*. If constant discounting does not hold, then characterizing one's time preference requires the specification of an entire discount *function*. Constant discounting implies that a person's intertemporal preferences are *time-consistent*, which means that later preferences “confirm” earlier preferences. Formally, a person's preferences are time-consistent if, for any two consumption profiles $(c_t, \ldots, c_T)$ and $(c'_t, \ldots, c'_T)$, with $c_t = c'_t$, $U^t(c_t, c_{t+1}, \ldots, c_T) \geq U^t(c'_t, c'_{t+1}, \ldots, c'_T)$ if and only if $U^{t+1}(c_{t+1}, \ldots, c_T) \geq U^{t+1}(c'_{t+1}, \ldots, c'_T)$.[7] For an interesting discussion that questions the normative validity of constant discounting see Albrecht and Weber (1995).

[6] An alternative but equivalent definition of constant discounting is that $D(k)/D(k + 1)$ is independent of k.

[7] Constant discounting implies time-consistent preferences only under the ancillary assumption of stationary discounting, for which the discount function $D(k)$ is the same in all periods. As a counterexample, if the period-t discount function is

$$D_t(k) = \left(\frac{1}{1+\rho}\right)^k$$

Diminishing Marginal Utility and Positive Time Preference

While not core features of the DU model, virtually all analyses of intertemporal choice assume both diminishing marginal utility (that the instantaneous utility function $u(c_t)$ is concave) and positive time preference (that the discount rate ρ is positive).[8] These two assumptions create opposing forces in intertemporal choice: diminishing marginal utility motivates a person to spread consumption over time, while positive time preference motivates a person to concentrate consumption in the present.

Since people do, in fact, spread consumption over time, the assumption of diminishing marginal utility (or some other property that has the same effect) seems strongly justified. The assumption of positive time preference, however, is more questionable. Several researchers have argued for positive time preference on logical grounds (Hirshleifer 1970; Koopmans 1960; Koopmans, Diamond, and Williamson 1964; Olson and Bailey 1981). The gist of their arguments is that a zero or negative time preference, combined with a positive real rate of return on saving, would command the infinite deferral of all consumption.[9] But this conclusion assumes, unrealistically, that individuals have infinite life spans and linear (or weakly concave) utility functions. Nevertheless, in econometric analyses of savings and intertemporal substitution, positive time preference is sometimes treated as an identifying restriction whose violation is interpreted as evidence of misspecification.

The most compelling argument supporting the logic of positive time preference was made by Derek Parfit (1971, 1976, 1982), who contends that there is no enduring self or "I" over time to which all future utility can be ascribed, and that a diminution in psychological connections gives our descendent future selves the status of other people—making that utility less than fully "ours" and giving us a reason to count it less.[10]

---

while the period-$t + 1$ discount function is

$$D_{t+1}(k) = \left(\frac{1}{1+\rho'}\right)^k$$

for some $\rho' \neq \rho$, then the person exhibits constant discounting at both dates t and $t + 1$, but nonetheless has time-inconsistent preferences.

[8] Discounting is not inherent to the DU model, because the model could be applied with $\rho \leq 0$. The inclusion of ρ in the model, however, strongly implies that it may take a value other than zero, and the name *discount rate* certainly suggests that it is greater than zero.

[9] In the context of intergenerational choice, Koopmans (1967) called this result the *paradox of the indefinitely postponed splurge*. See also Arrow (1983); Chakravarty (1962); and Solow (1974).

[10] As noted by Frederick (2002), there is much disagreement about the nature of Parfit's claim. In her review of the philosophical literature, Jennifer Whiting (1986, 549) identifies four different interpretations: the *strong absolute claim*: that it is irrational for someone to care about their future welfare; the *weak absolute claim*: that there is no rational requirement to care about one's future welfare; the *strong comparative claim*: that it is irrational to care more about one's own future welfare than about the welfare of any other person; and the *weak comparative claim*: that one is not rationally required to care more about his or her future welfare than about the welfare of any other person. We believe that all of these interpretations are too strong, and that Parfit endorses only a weaker version of the weak absolute claim. That is, he claims only that one is not rationally required to care about

> We care less about our further future . . . because we know that less of what we are now—less, say, of our present hopes or plans, loves or ideals—will survive into the further future . . . [if] what matters holds to a lesser degree, it cannot be irrational to care less. (Parfit 1971, p. 99)

Parfit's claims are normative, not descriptive. He is not attempting to explain or predict people's intertemporal choices, but is arguing that conclusions about the rationality of time preference must be grounded in a correct view of personal identity. If this is the only compelling normative rationale for time discounting, however, it would be instructive to test for a positive relation between observed time discounting and changing identity. Frederick (1999) conducted the only study of this type, and found no relation between monetary discount rates (as imputed from procedures such as "I would be indifferent between $100 tomorrow and $——— in five years") and self-perceived stability of identity (as defined by the following similarity ratings: "Compared to now, how similar were you five years ago [will you be five years from now]?"), nor did he find any relation between such monetary discount rates and the presumed correlates of identity stability (for example, the extent to which people agree with the statement "I am still embarrassed by stupid things I did a long time ago").

Discounted Utility Anomalies

Over the past two decades, empirical research on intertemporal choice has documented various inadequacies of the DU model as a descriptive model of behavior. First, empirically observed discount rates are not constant over time, but appear to decline—a pattern often referred to as hyperbolic discounting. Furthermore, even for a given delay, discount rates vary across different types of intertemporal choices: gains are discounted more than losses, small amounts more than large amounts, and explicit sequences of multiple outcomes are discounted differently than outcomes considered singly.

Hyperbolic Discounting

The best documented DU anomaly is hyperbolic discounting. The term *hyperbolic discounting* is often used to mean that a person has a declining rate of time preference (in our notation, ρ_n is declining in n), and we adopt this meaning here. Several results are usually interpreted as evidence for hyperbolic discounting. First, when subjects are asked to compare a smaller-sooner reward to a larger-later reward (to be discussed), the implicit discount rate over longer time horizons is lower than the implicit discount rate over shorter time horizons. For example, Thaler (1981) asked subjects to specify the amount of money they would require

one's future welfare to a degree that exceeds the degree of psychological connectedness that obtains between one's current self and one's future self.

in 1 month, 1 year, and 10 years to make them indifferent to receiving $15 now. The median responses—$20, $50, $100—imply an average (annual) discount rate of 345% over a one-month horizon, 120% over a 1-year horizon, and 19% over a 10-year horizon.[11] Other researchers have found a similar pattern (Benzion, Rapoport, and Yagil 1989; Chapman 1996; Chapman and Elstein 1995; Pender 1996; Redelmeier and Heller 1993).

Second, when mathematical functions are explicitly fit to such data, a hyperbolic functional form, which imposes declining discount rates, fits the data better than the exponential functional form, which imposes constant discount rates (Kirby 1997; Kirby and Marakovic 1995; Myerson and Green 1995; Rachlin, Raineri, and Cross 1991).[12]

Third, researchers have shown that preferences between two delayed rewards can reverse in favor of the more proximate reward as the time to both rewards diminishes—for example, someone may prefer $110 in 31 days over $100 in 30 days, but also prefer $100 now over $110 tomorrow. Such "preference reversals" have been observed both in humans (Green, Fristoe, and Myerson 1994; Kirby and Herrnstein 1995; Millar and Navarick 1984; Solnick et al. 1980) and in pigeons (Ainslie and Herrnstein 1981; Green et al. 1981).[13]

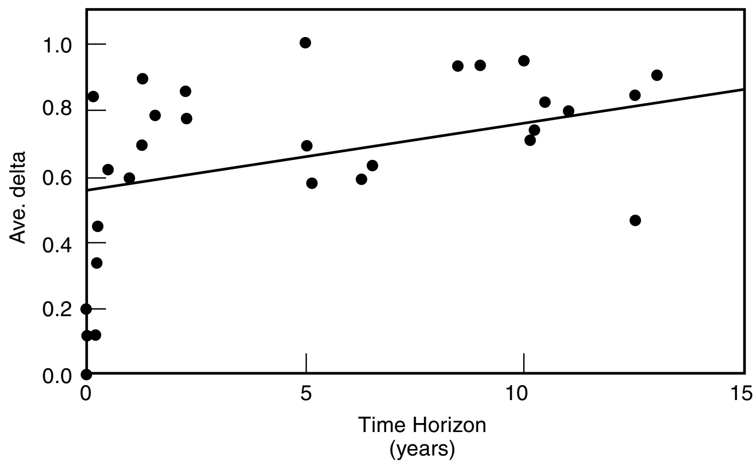
Fourth, the pattern of declining discount rates suggested by these studies is also evident *across* studies. Figure 6.1a plots the average estimated discount factor (= 1/(1 + discount rate)) from each of these studies against the average time horizon for that study.[14] As the regression line reflects, the estimated discount factor increases with the time horizon, which means that the discount rate declines. We note, however, that after excluding studies with very short time horizons (one year or less) from the analysis (see figure 6.1b), there is no evidence that discount

[11] That is, $\$15 = \$20 \times (e^{-(3.45)(1/12)}) = \$50 \times (e^{-(1.20)(1)}) = \$100 \times (e^{-(0.19)(10)})$. While most empirical studies report average discount rates over a given horizon, it is sometimes more useful to discuss average "per-period" discount rates. Framed in these terms, Thaler's results imply an average (annual) discount rate of 345 percent between now and one month from now, 100% between 1 month from now and 1 year from now, and 7.7% between 1 year from now and 10 years from now. That is, $\$15 = \$20{*}(e^{-(3.45)(1/12)}) = \$50{*}(e^{-(3.45)(1/12)}\ e^{-(1.00)(11/12)}) = \$100{*}(e^{-(3.45)(1/12)}\ e^{-(1.00)(11/12)}\ e^{-(0.077)(9)})$.

[12] Several hyperbolic functional forms have been proposed: Ainslie (1975) suggested the function $D(t) = 1/t$, Herrnstein (1981) and Mazur (1987) suggested $D(t) = 1/(1 + \alpha t)$, and Loewenstein and Prelec (1992) suggested $D(t) = 1/(1 + \alpha t)^{\beta/\alpha}$.

[13] These studies all demonstrate preference reversals in the synchronic sense—subjects simultaneously prefer $100 now over $110 tomorrow and prefer $110 in 31 days over $100 in 30 days, which is consistent with hyperbolic discounting. Yet there seems to be an implicit belief that such preference reversals would also hold in the diachronic sense—that if subjects who currently prefer $110 in 31 days over $100 in 30 days were brought back to the lab 30 days later, they would prefer $100 at that time over $110 one day later. Under the assumption of stationary discounting (as discussed earlier), synchronic preference reversals imply diachronic preference reversals. To the extent that subjects anticipate diachronic reversals and want to avoid them, evidence of a preference for commitment could also be interpreted as evidence for hyperbolic discounting (to be discussed).

[14] In some cases, the discount rates were computed from the median respondent. In other cases, the mean discount rate was used.



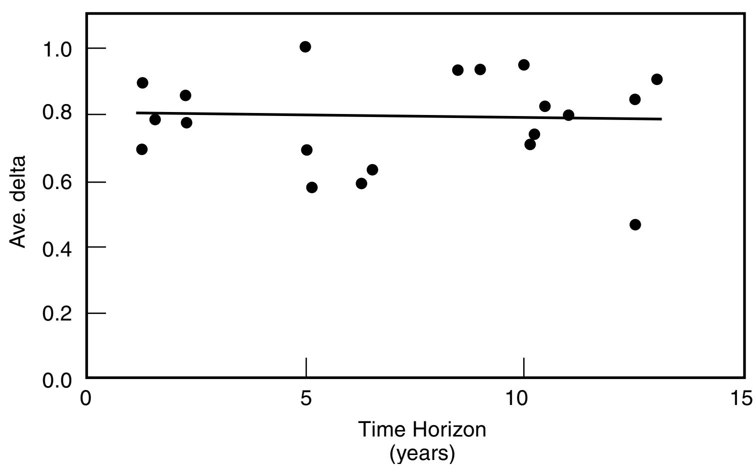


FIGURE 6.1a–b Discount factor as a function of time horizon (a) all studies. (b) studies with ave. horizons > 1 year.
Source: Frederick, Loewenstein, and O'Donoghue (2002).

rates continue to decline. In fact, after excluding the studies with short time horizons, the correlation between time horizon and discount factor is almost exactly zero (−0.0026).

Although the collective evidence outlined here seems overwhelmingly to support hyperbolic discounting, a recent study by Read (2001) points out that the most common type of evidence—the finding that implicit discount rates decrease with the time horizon—could also be explained by "subadditive discounting," which means that the total amount of discounting over a temporal interval increases as the interval is more finely partitioned.[15] To demonstrate subadditive discounting and distinguish it from hyperbolic discounting, Read elicited discount rates for a 2-year (24-month) interval and for its 3 constituent intervals, an 8-month interval beginning at the same time, an 8-month interval beginning 8 months later, and an 8-month interval beginning 16 months later. He found that the average discount rate for the 24-month interval was lower than the compounded average discount rate over the 3 8-month subintervals—a result predicted by subadditive discounting but not predicted by hyperbolic discounting (or any type of discount function, for that matter). Moreover, there was no evidence that discount rates declined with time, as the discount rates for the 3 8-month intervals were approximately equal. Similar empirical results were found earlier by Holcomb and Nelson (1992), although they did not interpret their results the same way.

If Read is correct about subadditive discounting, its main implication for economic applications may be to provide an alternative psychological underpinning for using a hyperbolic discount function, because most intertemporal decisions are based primarily on discounting from the present.[16]

Other DU Anomalies

The DU model not only dictates that the discount rate should be constant for all time periods, it also assumes that the discount rate should be the same for all types of goods and all categories of intertemporal decisions. There are several empirical regularities that appear to contradict this assumption, namely: gains are

[15] Read's proposal that discounting is subadditive is compatible with analogous results in other domains. For example, Tversky and Koehler (1994) found that the total probability assigned to an event increases the more finely the event is partitioned—for example, the probability of "death by accident" is judged to be more likely if one separately elicits the probability of "death by fire," "death by drowning," "death by falling," and so on.

[16] A few studies have actually found *increasing* discount rates. Frederick (1999) asked 228 respondents to imagine that they worked at a job that consisted of both pleasant work ("good days") and unpleasant work ("bad days") and to equate the attractiveness of having additional good days this year or in a future year. On average, respondents were indifferent between twenty extra good days this year, twenty-one the following year, or forty in five years, implying a 1-year discount rate of 5% and a 5-year discount rate of 15%. A possible explanation is that a desire for improvement is evoked more strongly for 2 successive years (this year and next) than for 2 separated years (this year and 5 years hence). Rubinstein (2000) asked students in a political science class to choose, between the following two payment sequences:

	March 1	June 1	Sept. 1	Nov. 1
A:	$997	$997	$997	$997
	April 1	July 1	Oct. 1	Dec. 1
B:	$1,000	$1,000	$1,000	$1,000

Then, two weeks later, he asked them to choose between $997 on November 1 and $1,000 on December 1. Fifty-four percent of respondents preferred $997 in November to $1,000 in December, but only 34% preferred sequence A to sequence B. These two results suggest increasing discount rates. To explain them, Rubinstein speculated that the three more proximate additional elements may have masked the differences in the timing of the sequence of dated amounts, while making the differences in amounts more salient.

discounted more than losses; small amounts are discounted more than large amounts; greater discounting is shown to avoid delay of a good than to expedite its receipt; in choices over sequences of outcomes, improving sequences are often preferred to declining sequences though positive time preference dictates the opposite; and in choices over sequences, violations of independence are pervasive, and people seem to prefer spreading consumption over time in a way that diminishing marginal utility alone cannot explain.

THE "SIGN EFFECT" (GAINS ARE DISCOUNTED MORE THAN LOSSES)

Many studies have concluded that gains are discounted at a higher rate than losses. For instance, Thaler (1981) asked subjects to imagine they had received a traffic ticket that could be paid either now or later and to state how much they would be willing to pay if payment could be delayed (by three months, one year, or three years). The discount rates imputed from these answers were much lower than the discount rates imputed from comparable questions about monetary gains. This pattern is prevalent in the literature. Indeed, in many studies, a substantial proportion of subjects prefer to incur a loss immediately rather than delay it (Benzion, Rapoport, and Yagil 1989; Loewenstein 1987; MacKeigan et al. 1993; Mischel, Grusec, and Masters 1969; Redelmeier and Heller 1993; Yates and Watts 1975).

THE "MAGNITUDE EFFECT" (SMALL OUTCOMES ARE DISCOUNTED MORE THAN LARGE ONES)

Most studies that vary outcome size have found that large outcomes are discounted at a lower rate than small ones (Ainslie and Haendel 1983; Benzion, Rapoport, and Yagil 1989; Green, Fristoe, and Myerson 1994; Green, Fry, and Myerson 1994; Holcomb and Nelson 1992; Kirby 1997; Kirby and Marakovic 1995; Kirby, Petry, and Bickel 1999; Loewenstein 1987; Raineri and Rachlin 1993; Shelley 1993; Thaler 1981). In Thaler's (1981) study, for example, respondents were, on average, indifferent between $15 immediately and $60 in a year, $250 immediately and $350 in a year, and $3,000 immediately and $4,000 in a year, implying discount rates of 139%, 34%, and 29%, respectively.

THE "DELAY-SPEEDUP" ASYMMETRY

Loewenstein (1988) demonstrated that imputed discount rates can be dramatically affected by whether the change in delivery time of an outcome is framed as an acceleration or a delay from some temporal reference point. For example, respondents who didn't expect to receive a VCR for another year would pay an average of $54 to receive it immediately, but those who thought they would receive it immediately demanded an average of $126 to delay its receipt by a year. Benzion, Rapoport, and Yagil (1989) and Shelley (1993) replicated Loewenstein's findings for losses as well as gains (respondents demanded more to expedite payment than they would pay to delay it).

PREFERENCE FOR IMPROVING SEQUENCES

In studies of discounting that involve choices between two outcomes—for example, X at τ versus Y at τ'—positive discounting is the norm. Research examining preferences over *sequences* of outcomes, however, has generally found that people prefer improving sequences to declining sequences (for an overview see Ariely and Carmon, in press; Frederick and Loewenstein 2002; Loewenstein and Prelec 1993). For example, Loewenstein and Sicherman (1991) found that, for an otherwise identical job, most subjects prefer an increasing wage profile to a declining or flat one (see also Frank 1993). Hsee, Abelson, and Salovey (1991) found that an increasing salary sequence was rated as highly as a decreasing sequence that conferred much more money. Varey and Kahneman (1992) found that subjects strongly preferred streams of decreasing discomfort to streams of increasing discomfort, even when the overall sum of discomfort over the interval was otherwise identical. Loewenstein and Prelec (1993) found that respondents who chose between sequences of two or more events (for example, dinners or vacation trips) on consecutive weekends or consecutive months generally preferred to save the better thing for last. Chapman (2000) presented respondents with hypothetical sequences of headache pain that were matched in terms of total pain that either gradually lessened or gradually increased with time. Sequence durations included one hour, one day, one month, one year, five years, and twenty years. For all sequence durations, the vast majority (from 82 to 92%) of subjects preferred the sequence of pain that lessened over time (see also Ross and Simonson 1991).

VIOLATIONS OF INDEPENDENCE AND PREFERENCE FOR SPREAD

The research on preferences over sequences also reveals strong violations of independence. Consider the following pair of questions from Loewenstein and Prelec (1993):

> Imagine that over the next five weekends you must decide how to spend your Saturday nights. From each pair of sequences of dinners below, circle the one you would prefer. "Fancy French" refers to a dinner at a fancy French Restaurant. "Fancy lobster" refers to an exquisite lobster dinner at a four-star restaurant. Ignore scheduling considerations (e.g., your current plans).

Options	*First weekend*	*Second weekend*	*Third weekend*	*Fourth weekend*	*Fifth weekend*	
A	Fancy French	Eat at home	Eat at home	Eat at home	Eat at home	[11%]
B	Eat at home	Eat at home	Fancy French	Eat at home	Eat at home	[89%]

Options	*First weekend*	*Second weekend*	*Third weekend*	*Fourth weekend*	*Fifth weekend*	
C	Fancy French	Eat at home	Eat at home	Eat at home	Fancy lobster	[49%]
D	Eat at home	Eat at home	Fancy French	Eat at home	Fancy lobster	[51%]

As discussed earlier, consumption independence implies that preferences between two consumption profiles should not be affected by the nature of the consumption in periods in which consumption is identical in the two profiles. Thus, anyone preferring profile B to profile A (which share the fifth period "Eat at home") should also prefer profile D to profile C (which share the fifth period "Fancy lobster"). As the data reveal, however, many respondents violated this prediction, preferring the fancy French dinner on the third weekend, if that was the only fancy dinner in the profile, but preferring the fancy French dinner on the first weekend if the profile contained another fancy dinner. This result could be explained by the simple desire to spread consumption over time—which, in this context, violates the dubious assumption of independence that the DU model entails.

Loewenstein and Prelec (1993) provide further evidence of such a preference for spread. Subjects were asked to imagine that they were given two coupons for fancy ($100) restaurant dinners, and were asked to indicate when they would use them, ignoring considerations such as holidays, birthdays, and such. Subjects were told either that "you can use the coupons at any time between today and two years from today" or were told nothing about any constraints. Subjects in the 2-year constraint condition actually scheduled both dinners at a *later* time than those who faced no explicit constraint—they delayed the first dinner for 8 weeks (rather than 3) and the second dinner for 31 weeks (rather than 13). This counterintuitive result can be explained in terms of a preference for spread if the explicit two-year interval was greater than the implicit time horizon of subjects in the unconstrained group.

Are These "Anomalies" Mistakes?

In other domains of judgment and choice, many of the famous "effects" that have been documented are regarded as errors by the people who commit them. For example, in the "conjunction fallacy" discovered by Tversky and Kahneman (1983), many people will—with some reflection—recognize that a conjunction cannot be more likely than one of its constituents (for example, that it *can't* be more likely for Linda to be a feminist bank teller than for her to be "just" a bank teller). In contrast, the patterns of preferences that are regarded as "anomalies" in the context of the DU model do not necessarily violate any standard or principle that people believe they should uphold. Even when the choice pattern is pointed out to people, they do not regard themselves as having made a mistake (and probably have not made one!). For example, there is no compelling logic that dictates that one who prefers to delay a French dinner should also prefer to do so when that French dinner will be closely followed by a lobster dinner.

Indeed, it is unclear whether *any* of the DU "anomalies" should be regarded as mistakes. Frederick and Read (2002) found evidence that the magnitude effect is *more* pronounced when subjects evaluate both "small" and "large" amounts than when they evaluate either one. Specifically, the difference in the discount rates between a small amount ($10) and a large amount ($1,000) was larger when the two judgments were made in close succession than when made separately. Analogous results were obtained for the sign effect as the differences in discount rates between gains and losses were slightly larger in a within-subjects design, where respondents evaluated delayed gains and delayed losses, than in a between-subjects design, where they evaluate only gains or only losses. Since respondents did not attempt to coordinate their responses to conform to DU's postulates when they evaluated rewards of different sizes, it suggests that they consider the different discount rates to be normatively appropriate. Similarly, even after Loewenstein and Sicherman (1991) informed respondents that a decreasing wage profile ($27,000, $26,000, . . . $23,000) would (via appropriate saving and investing) permit strictly more consumption in every period than the corresponding increasing wage profile with an equivalent nominal total ($23,000, $24,000, . . . $27,000), respondents still preferred the increasing sequence. Perhaps they suspected that they could not exercise the required self-control to maintain their desired *consumption* sequence, or felt a general leeriness about the significance of a declining wage, either of which could justify that choice. As these examples illustrate, many DU "anomalies" exist as "anomalies" only by reference to a model that was constructed without regard to its descriptive validity, and which has no compelling normative basis.

Alternative Models

In response to the anomalies just enumerated, and other intertemporal-choice phenomena that are inconsistent with the DU model, a variety of alternate theoretical models have been developed. Some models attempt to achieve greater descriptive realism by relaxing the assumption of constant discounting. Other models incorporate additional considerations into the instantaneous utility function, such as the utility from anticipation. Still others depart from the DU model more radically, by including, for instance, systematic mispredictions of future utility.

Models of Hyperbolic Discounting

In the economics literature, Strotz was the first to consider alternatives to exponential discounting, seeing "no reason why an individual should have such a special discount function" (1955–56, p. 172). Moreover, Strotz recognized that for any discount function other than exponential, a person would have time-inconsistent preferences.[17] He proposed two strategies that might be employed by a person who foresees how her preferences will change over time: the "strategy of precommitment" (wherein she commits to some plan of action) and the "strategy of consistent planning" (wherein she chooses her behavior ignoring plans that she knows her future selves will not carry out).[18] While Strotz did not posit any

[17] Strotz implicitly assumes stationary discounting.

[18] Building on Strotz's strategy of consistent planning, some researchers have addressed the question of whether a consistent path exists for general nonexponential discount functions. See in particular Pollak (1968); Peleg and Yaari (1973); and Goldman (1980).

specific alternative functional forms, he did suggest that "special attention" be given to the case of declining discount rates.

Motivated by the evidence discussed earlier, there has been a recent surge of interest among economists in the implications of declining discount rates (beginning with Laibson 1994, 1997). This literature has used a particularly simple functional form that captures the essence of hyperbolic discounting:

$$D(k) = \begin{cases} 1 & \text{if } h = 0 \\ \beta\delta^k & \text{if } k > 0. \end{cases}$$

This functional form was first introduced by Phelps and Pollak (1968) to study intergenerational altruism, and was first applied to individual decision making by Elster (1979). It assumes that the per-period discount rate between now and the next period is $(1 - \beta\delta)/\beta\delta$ whereas the per-period discount rate between any two future periods is

$$\frac{1-\delta}{\delta} < \frac{1-\beta\delta}{\beta\delta}.$$

Hence, this (β,δ) formulation assumes a declining discount rate between this period and next, but a constant discount rate thereafter. The (β,δ) formulation is highly tractable, and captures many of the *qualitative* implications of hyperbolic discounting.

Laibson and his collaborators have used the (β,δ) formulation to explore the implications of hyperbolic discounting for consumption-saving behavior. Hyperbolic discounting leads one to consume more than one would like to from a prior perspective (or, equivalently, to undersave). Laibson (1997) explores the role of illiquid assets, such as housing, as an imperfect commitment technology, emphasizing how one could limit overconsumption by tying up one's wealth in illiquid assets. Laibson (1998) explores consumption-saving decisions in a world without illiquid assets (or any other commitment technology). These papers describe how hyperbolic discounting might explain some stylized empirical facts, such as the excess comovement of income and consumption, the existence of asset-specific marginal propensities to consume, low levels of precautionary savings, and the correlation of measured levels of patience with age, income, and wealth. Laibson, Repetto, and Tobacman (1998), and Angeletos and colleagues (2001) calibrate models of consumption-saving decisions, using both exponential discounting and (β,δ) hyperbolic discounting. By comparing simulated data to real-world data, they demonstrate how hyperbolic discounting can better explain a variety of empirical observations in the consumption-saving literature. In particular, Angeletos and colleagues (2001) describe how hyperbolic discounting can explain the coexistence of high preretirement wealth, low liquid asset holdings (relative to income levels and illiquid asset holdings), and high credit-card debt.

Fischer (1999) and O'Donoghue and Rabin (1999c, 2001) have applied (β,δ) preferences to procrastination, where hyperbolic discounting leads a person to put off an onerous activity more than she would like to from a prior perspective.[19] O'Donoghue and Rabin (1999c) examine the implications of hyperbolic discounting for contracting when a principal is concerned with combating procrastination by an agent. They show how incentive schemes with "deadlines" may be a useful screening device to distinguish efficient delay from inefficient procrastination. O'Donoghue and Rabin (2001) explore procrastination when a person must not only choose *when* to complete a task, but also *which* task to complete. They show that a person might never carry out a very easy and very good option because they continually *plan* to carry out an even better but more onerous option. For instance, a person might never take half an hour to straighten the shelves in her garage because she persistently plans to take an entire day to do a major cleanup of the entire garage. Extending this logic, they show that providing people with new options might make procrastination more likely. If the person's only option were to straighten the shelves, she might do it in a timely manner; but if the person can either straighten the shelves or do the major cleanup, she now may do nothing. O'Donoghue and Rabin (1999d) apply this logic to retirement planning.

O'Donoghue and Rabin (1999a, 2000), Gruber and Koszegi (2000), and Carrillo (1999) have applied (β,δ) preferences to addiction. These researchers describe how hyperbolic discounting can lead people to overconsume harmful addictive products, and examine the degree of harm caused by such overconsumption. Carrillo and Mariotti (2000) and Benabou and Tirole (2000) have examined how (β,δ) preferences might influence a person's decision to acquire information. If, for example, one is deciding whether to embark on a specific research agenda, one may have the option to get feedback from colleagues about its likely fruitfulness. The standard economic model implies that people should always choose to acquire this information if it is free. Carrillo and Mariotti show, however, that hyperbolic discounting can lead to "strategic ignorance"—a person with hyperbolic discounting who is worried about withdrawing from an advantageous course of action when the costs become imminent might choose not to acquire free information if doing so increases the risk of bailing out.

Self-Awareness

A person with time-inconsistent preferences may or may not be aware that his or her preferences will change over time. Strotz (1955–56) and Pollak (1968) discussed two extreme alternatives. At one extreme, a person could be completely "naive" and believe that her future preferences will be identical to her current preferences. At the other extreme, a person could be completely "sophisticated" and correctly predict how his or her preferences will change over time. While casual observation and introspection suggest that people lie somewhere between these two extremes, behavioral evidence regarding the degree of awareness is quite limited.

[19] While not framed in terms of hyperbolic discounting, Akerlof's (1991) model of procrastination is formally equivalent to a hyperbolic model.

One way to identify sophistication is to look for evidence of commitment. Someone who suspects that his or her preferences will change over time might take steps to eliminate an inferior option that might tempt one later. For example, someone who currently prefers $110 in 31 days to $100 in 30 days but who suspects that in a month she will prefer $100 immediately to $110 tomorrow, might attempt to eliminate the $100 reward from the later choice set, and thereby bind herself *now* to receive the $110 reward in 31 days. Real-world examples of commitment include "Christmas clubs" or "fat farms."

Perhaps the best empirical demonstration of a preference for commitment was conducted by Ariely and Wertenbroch (2002). In that study, MIT executive-education students had to write three short papers for a class and were assigned to one of two experimental conditions. In one condition, deadlines for the three papers were imposed by the instructor and were evenly spaced across the semester. In the other condition, each student was allowed to set his or her own deadlines for each of the three papers. In both conditions, the penalty for delay was 1 percent per day late, regardless of whether the deadline was externally imposed or self-imposed. Although students in the free-choice condition could have made all three papers due at the end of the semester, many in fact did choose to impose deadlines on themselves, suggesting that they appreciated the value of commitment. Few students chose evenly spaced deadlines, however, and those who did not performed worse in the course than those with evenly spaced deadlines (whether externally imposed or self-imposed).[20]

O'Donoghue and Rabin (1999b) examine how people's behaviors depend on their sophistication about their own time inconsistency. Some behaviors, such as using illiquid assets for commitment, require some degree of sophistication. Other behaviors, such as overconsumption or procrastination, are more robust to the degree of awareness, though the degree of misbehavior may depend on the degree of sophistication. To understand such effects, O'Donoghue and Rabin (2001) introduce a formal model of *partial naïvete*, in which a person is aware that he or she will have future self-control problems but under-estimates their magnitude. They show that severe procrastination cannot occur under complete sophistication, but can arise if the person is only a little naïve. (For more discussion on self-awareness see O'Donoghue and Rabin, chap. 7 in this volume.)

The degree of sophistication versus naïvete has important implications for public policy. If people are sufficiently sophisticated about their own self-control problems, providing commitment devices may be beneficial. If people are naïve, however, policies might be better aimed at either educating people about loss of control (making them more sophisticated), or providing incentives for people to use commitment devices, even if they don't recognize the need for them.

[20] A similar "natural" experiment was recently conducted by the Economic and Social Research Council of Great Britain. They recently eliminated submission deadlines and now accept grant proposals on a "rolling" basis (though they are still reviewed only periodically). In response to this policy change, submissions have actually declined by 15 to 20% (direct correspondence with Chris Caswill at ESRC).

Models That Enrich the Instantaneous-Utility Function

Many discounting anomalies, especially those discussed earlier, can be understood as a misspecification of the instantaneous-utility function. Similarly, many of the confounds discussed in the section on measuring time discounting are caused by researchers attributing to the discount rate aspects of preference that are more appropriately considered as arguments in the instantaneous utility function. As a result, alternative models of intertemporal choice have been advanced that add additional arguments, such as utility from anticipation, to the instantaneous-utility function.

HABIT-FORMATION MODELS

James Duesenberry (1952) was the first economist to propose the idea of "habit formation"—that the utility from current consumption ("tastes") can be affected by the level of past consumption. This idea was more formally developed by Pollak (1970) and Ryder and Heal (1973). In habit-formation models, the period-τ instantaneous utility function takes the form $u\ (c_\tau,\ c_{\tau-1},\ c_{\tau-2}, \ldots)$ where $\partial^2 u/\partial c_\tau \partial c_{\tau'} > 0$ for $\tau' < \tau$. For simplicity, most such models assume that all effects of past consumption for current utility enter through a state variable. That is, they assume that period-τ instantaneous-utility function takes the form $u(c_\tau;\ z_\tau)$, where z_τ is a state variable that is increasing in past consumption and $\partial^2/\partial c_\tau \partial z_{\tau'} > 0$. Both Pollak (1970) and Ryder and Heal (1973) assume that z_τ is the exponentially weighted sum of past consumption, or

$$z_\tau = \sum_{i=1}^{\infty} \tilde{a}^i c_{\tau-i}.$$

Although habit formation is often said to induce a preference for an increasing consumption profile, it can, under some circumstances, lead a person to prefer a decreasing or even nonmonotonic consumption profile. The direction of the effect depends on things such as how much one has already consumed (as reflected in the initial habit stock), and perhaps most important, whether current consumption increases or decreases future utility.

In recent years, habit-formation models have been used to analyze a variety of phenomena. Becker and Murphy (1988) use a habit-formation model to study addictive activities, and in particular to examine the effects of past and future prices on the current consumption of addictive products.[21] Habit formation can help explain asset-pricing anomalies such as the equity-premium puzzle (Abel 1990; Campbell and Cochrane 1999; Constantinides 1990). Incorporating habit formation into business-cycle models can improve their ability to explain movements in asset prices (Jermann 1998; Boldrin, Christiano, and Fisher 2001). Some recent

[21] For rational-choice models building on Becker and Murphy's framework see Orphanides and Zervos (1995); Wang (1997); and Suranovic, Gold-farb, and Leonard (1999). For addiction models that incorporate hyperbolic discounting see O'Donoghue and Rabin (1999a, 2000); Gruber and Koszegi (2000); and Carrillo (1999).

papers have shown that habit formation may help explain other empirical puzzles in macroeconomics as well. Whereas standard growth models assume that high saving rates cause high growth, recent evidence suggests that the causality can run in the opposite direction. Carroll, Overland, and Weil (2000) show that, under conditions of habit formation, high growth rates can cause people to save more. Fuhrer (2000) shows how habit formation might explain the recent finding that aggregate spending tends to have a gradual "hump-shaped" response to various shocks. The key feature of habit formation that drives many of these results is that, after a shock, consumption adjustment is sluggish in the short term but not in the long term.

REFERENCE-POINT MODELS

Closely related to, but conceptually distinct from, habit-formation models are models of reference-dependent utility, which incorporate ideas from prospect theory (Kahneman and Tversky 1979; Tversky and Kahneman 1991). According to prospect theory, outcomes are evaluated using a value function defined over departures from a reference point—in our notation, the period-τ instantaneous utility function takes the form $u(c_\tau, r_\tau) = v(c_\tau - r_\tau)$. The reference point, r_τ, might depend on past consumption, expectations, social comparison, status quo, and such. A second feature of prospect theory is that the value function exhibits *loss-aversion*—negative departures from one's reference consumption level decrease utility by a greater amount than positive departures increase it. A third feature of prospect theory is that the value function exhibits *diminishing sensitivity* for both gains and losses, which means that the value function is concave over gains and convex over losses.[22]

Loewenstein and Prelec (1992) applied a specialized version of such a value function to intertemporal choice to explain the magnitude effect, the sign effect, and the delay-speedup asymmetry. They show that if the elasticity of the value function is increasing in the magnitude of outcomes, people will discount smaller magnitudes more than larger magnitudes. Intuitively, the elasticity condition captures the insight that people are responsive to both differences and ratios of reward amounts. It implies that someone who is indifferent between, say, \$10 now and \$20 in a year should prefer \$200 in a year over \$100 now because the larger rewards have a greater difference (and the same ratio). Consequently, even if one's time preference is actually constant across outcomes, a person will be more willing to wait for a fixed proportional increment when rewards are larger and, thus, one's imputed discount rate will be smaller for larger outcomes. Similarly, if the value function for losses is more elastic than the value function for gains, then people will discount gains more than losses. Finally, such a model helps explain

[22] Reference-point models sometimes assume a direct effect of the consumption level or reference level, so that $u(c_\tau, r_\tau) = v(c_\tau - r_\tau) + w(c_\tau)$ or $u(c_\tau, r_\tau) = v(c_\tau - r_\tau) + w(r_\tau)$. Some habit-formation models could be interpreted as reference-point models, where the state variable z_τ is the reference point. Indeed, many habit-formation models, such as Pollak (1970) and Constantinides (1990), assume instantaneous utility functions of the form $u(c_\tau - z_\tau)$, although they typically assume neither loss aversion nor diminishing sensitivity.

the delay-speedup asymmetry (Loewenstein 1988). Shifting consumption in any direction is made less desirable by loss-aversion, since one loses consumption in one period and gains it in another. When delaying consumption, loss-aversion reinforces time discounting, creating a powerful aversion to delay. When expediting consumption, loss-aversion opposes time discounting, reducing the desirability of speedup (and occasionally even causing an aversion to it).

Using a reference-dependent model that assumes loss aversion in consumption, Bowman, Minehart, and Rabin (1999) predict that "news" about one's (stochastic) future income affects one's consumption growth differently than the standard Permanent Income Hypothesis predicts. According to (the log-linear version of) the Permanent Income Hypothesis, changes in future income should not affect the rate of consumption growth. For example, if a person finds out that his or her permanent income will be lower than formerly thought, he or she would reduce consumption by, say, 10 percent in every period, leaving consumption growth unchanged. If, however, this person were loss-averse in current consumption, he or she would be unwilling to reduce this year's consumption by 10 percent—forcing that person to reduce future consumption by *more* than 10 percent, and thereby reducing the growth rate of consumption. Two studies by Shea (1995a, 1995b) support this prediction. Using both aggregate U.S. data and data from teachers' unions (in which wages are set one year in advance), Shea finds that consumption growth responds more strongly to future wage decreases than to future wage increases.

MODELS INCORPORATING UTILITY FROM ANTICIPATION

Some alternative models build on the notion of "anticipal" utility discussed by the elder and younger Jevons. If people derive pleasure not only from current consumption but also from anticipating future consumption, then current instantaneous utility will depend positively on future consumption—that is, the period-τ instantaneous utility function would take the form $u(c_\tau; c_{\tau+1}, c_{\tau+2}, \ldots)$ where $\partial u/\partial c_{\tau'} > 0$ for $\tau' > \tau$. Loewenstein (1987) advanced a formal model that assumes that a person's instantaneous utility is equal to the utility from consumption in that period plus some function of the discounted utility of consumption in future periods. Specifically, if we let $v(c)$ denote utility from actual consumption, and assume this is the same for all periods, then:

$$u(c_\tau; c_{\tau+1}, c_{\tau+2}, \ldots) = v(c_\tau) + \acute{a}[\tilde{a}v(c_{\tau+1}) + \tilde{a}^2 v(c_{\tau+2}) + \ldots] \quad \text{for some } \tilde{a} < 1.$$

Loewenstein describes how utility from anticipation may play a role in many DU anomalies. Because near-term consumption delivers only consumption utility whereas future consumption delivers both consumption utility and anticipatory utility, anticipatory utility provides a reason to prefer improvement and for getting unpleasant outcomes over with quickly instead of delaying them as discounting would predict. It provides a possible explanation for why people discount different goods at different rates, because utility from anticipation creates a downward bias on estimated discount rates, and this downward bias is larger for goods that create more anticipatory utility. If, for instance, dreading future bad outcomes is a stronger emotion than savoring future good outcomes, which seems highly plausible, then utility from anticipation would generate a sign effect.[23]

Finally, anticipatory utility gives rise to a form of time inconsistency that is quite different from that which arises from hyperbolic discounting. Instead of planning to do the farsighted thing (for example, save money) but subsequently doing the shortsighted thing (splurging), anticipatory utility can cause people to repeatedly plan to consume a good after some delay that permits pleasurable anticipation, but then to delay again for the same reason when the planned moment of consumption arrives.

Loewenstein's model of anticipatory utility applies to deterministic outcomes. In a recent paper, Caplin and Leahy (2001) point out that many anticipatory emotions, such as anxiety or suspense, are driven by uncertainty about the future, and they propose a new model that modifies expected-utility theory to incorporate such anticipatory emotions. They then show that incorporating anxiety into asset-pricing models may help explain the equity premium puzzle and the risk-free rate puzzle, because anxiety creates a taste for risk-free assets and an aversion to risky assets. Like Loewenstein, Caplin and Leahy emphasize how anticipatory utility can lead to time inconsistency. Koszegi (2001) also discusses some implications of anticipatory utility.

VISCERAL INFLUENCES

A final alternative model of the utility function incorporates "visceral" influences such as hunger, sexual desire, physical pain, cravings, and such. Loewenstein (1996, 2000b) argues that economics should take more seriously the implications of such transient fluctuations in tastes. Formally, visceral influences mean that the person's instantaneous utility function takes the form $c_t = c_t'$ where d_τ represents the vector of visceral states in period τ. Visceral states are (at least to some extent) endogenous—for example, one's current hunger depends on how much one has consumed in previous periods—and therefore lead to consumption interdependence.

Visceral influences have important implications for intertemporal choice because, by increasing the attractiveness of certain goods or activities, they can give rise to behaviors that look extremely impatient or even impulsive. Indeed, for every visceral influence, it is easy to think of one or more associated problems of self-control—hunger and dieting, sexual desire and various "heat-of-the-moment" behaviors, craving and drug addiction, and so on. Visceral influences provide an alternate account of the preference reversals that are typically attributed to hyperbolic time discounting, because the temporal proximity of a reward is one of the cues that can activate appetitive visceral states (see Laibson 2001; Loewenstein 1996). Other cues—such as spatial proximity, the presence of associated smells or sounds, or similarity in current setting to historical consumption sites—may also have such an effect. Thus, research on various types of cues may help to

[23] Waiting for undesirable outcomes is almost always unpleasant, but waiting for desirable outcomes is sometimes pleasurable and sometimes frustrating. Despite the manifest importance for intertemporal choice of these emotions associated with waiting, we are aware of no research that has sought to understand when waiting for desirable outcomes is pleasurable or aversive.

generate new predictions about the specific circumstances (other than temporal proximity) that can trigger myopic behavior.

The fact that visceral states are endogenous introduces issues of state-management (as discussed by Loewenstein [1999] and Laibson [2001] under the rubric of "cue management"). While the model (atleast the rational version of it) predicts that one would want oneself to use drugs if one were to experience a sufficiently strong craving, it also predicts that one might want to prevent ever experiencing such a strong craving. Hence, visceral influences can give rise to a preference for commitment in the sense that the person may want to avoid certain situations.

Visceral influences may do more than merely change the instantaneous utility function. First, evidence shows that people don't fully appreciate the effects of visceral influences, and hence may not react optimally to them (Loewenstein 1996, 1999, 2000b). When in a hot state, people tend to exaggerate how long the hot state will persist, and, when in a cold state, people tend to underestimate how much future visceral influences will affect their future behavior. Second, and perhaps more importantly, people often would "prefer" not to respond to an intense visceral factor such as rage, fear, or lust, even at the moment they are succumbing to its influence. A way to understand such effects is to apply the distinction proposed by Kahneman (1994) between "experienced utility," which reflects one's welfare, and "decision utility," which reflects the attractiveness of options as inferred from one's decisions. By increasing the decision utility of certain types of actions more than the experienced utility of those actions, visceral factors may drive a wedge between what people do and what makes them happy. Bernheim and Rangel (2001) propose a model of addiction framed in these terms.

More "Extreme" Alternative Perspectives

The alternative models discussed thus far modify the DU model by altering the discount function or adding additional arguments to the instantaneous utility function. The alternatives discussed next involve more radical departures from the DU model.

PROJECTION BIAS

In many of the alternative models of utility discussed thus far, the person's utility from consumption—her tastes—change over time. To properly make intertemporal decisions, one must correctly predict how one's tastes will change. Essentially all economic models of changing tastes assume (as economists typically do) that such predictions are correct—that people have "rational expectations." Loewenstein, O'Donoghue, and Rabin (2000), however, propose that, while people may anticipate the qualitative nature of their changing preferences, they tend to underestimate the magnitude of these changes—a systematic misprediction they label *projection bias*.

Loewenstein, O'Donoghue, and Rabin review a broad array of evidence that demonstrates the prevalence of projection bias, then model it formally. To illustrate their model, consider projection bias in the realm of habit formation. As discussed earlier, suppose the period-τ instantaneous utility function takes the form $u(c_\tau; z_\tau)$, where z_τ is a state variable that captures the effects of past consumption. Projection bias arises when a person whose current state is z_t must predict his or her future utility given future state z_τ. Projection bias implies that the person's prediction $\tilde{u}(c_\tau; z_\tau | z_t)$ will lie between his or her true future utility $u(c_\tau; z_\tau)$ and his or her utility given the person's current state $u(c_\tau; z_\tau)$. A particularly simple functional form is $\tilde{u}(c_\tau; z_\tau | z_t) = (1 - a)\, u(c_\tau; z_\tau) + \alpha u(c_\tau; z_t)$ for some $\alpha \in [0,1]$.

Projection bias may arise whenever tastes change over time, whether through habit formation, changing reference points, or changes in visceral states. It can have important behavioral and welfare implications. For instance, people may underappreciate the degree to which a present consumption splurge will raise their reference consumption level, and thereby decrease their enjoyment of more modest consumption levels in the future. When intertemporal choices are influenced by projection bias, estimates of time preference may be distorted.

MENTAL-ACCOUNTING MODELS

Some researchers have proposed that people do not treat all money as fungible, but instead assign different types of expenditures to different "mental accounts" (see Thaler 1999 for a recent overview). Such models can give rise to intertemporal behaviors that seem odd when viewed through the lens of the DU model. Thaler (1985), for instance, suggests that small amounts of money are coded as spending money, whereas larger amounts of money are coded as savings, and that a person is more willing to spend out of the former account. This accounting rule would predict that people will behave like spendthrifts for small purchases (for example, a new pair of shoes), but act more frugally when it comes to large purchases (for example, a new dining-room table).[24] Benartzi and Thaler (1995) suggest that people treat their financial portfolios as a mental account, and emphasize the importance of how often people "evaluate" this account. They argue that if people review their portfolios once a year or so, and if people experience joy or pain from any gains or losses, as assumed in Kahneman and Tversky's (1979) prospect theory, then such "myopic loss-aversion" represents a plausible explanation for the equity premium puzzle.

Prelec and Loewenstein (1998) propose another way in which mental accounting might influence intertemporal choice. They posit that payments for consumption confer immediate disutility or "pain of paying," and that people keep mental accounts that link the consumption of a particular item with the payments for it. They also assume that people engage in "prospective accounting." According to prospective accounting, when consuming, people think only about current and future payments; past payments don't cause pain of paying. Likewise, when paying,

[24] While it seems possible that this conceptualization could explain the magnitude effect as well, the magnitude effect is found for very "small" amounts (for example, between \$2 and \$20 in Ainslie and Haendel [1983]), and for very "large amounts" (for example, between \$10,000 and \$1,000,000 in Raineri and Rachlin [1993]). It seems highly unlikely that respondents would consistently code the lower amounts as spending and the higher amounts as savings across all of these studies.

the pain of paying is buffered only by thoughts of future, but not past, consumption. The model suggests that different ways of financing a purchase can lead to different decisions, even holding the net present value of payments constant. Similarly, people might have different financing preferences depending on the consumption item (for example, they should prefer to prepay for a vacation that is consumed all at once versus a new car that is consumed over many years). The model generates a strong preference for prepayment (except for durables), for getting paid after rather than before doing work, and for fixed-fee pricing schemes with zero marginal costs over pay-as-you-go schemes that tightly couple marginal payments to marginal consumption. The model also suggests that interindividual heterogeneity might arise from differences in the degree to which people experience the pain of paying rather than differences in time preference. On this view, the miser who eschews a fancy restaurant dinner is not doing so because he or she explicitly considers the delayed costs of the indulgence, but rather because enjoyment of the dinner would be diminished by the immediate pain of paying for it.

CHOICE BRACKETING

One important aspect of mental accounting is that a person makes at most a few choices at any one time, and generally ignores the relation between these choices and other past and future choices. Which choices are considered at the same time is a matter of what Read, Loewenstein, and Rabin (1999) label *choice bracketing*. Intertemporal choices, like other choices, can be influenced by the manner in which they are bracketed, because different bracketing can highlight different motives. To illustrate, consider the conflict between impatience and a preference for improvement over time. Loewenstein and Prelec (1993) demonstrate that the relative importance of these two motives can be altered by the way that choices are bracketed. They asked one group of subjects to choose between having dinner at a fine French restaurant in one month versus two months. Most subjects chose one month, presumably reflecting impatience. They then asked another group to choose between eating at home in one month followed by eating at the French restaurant in two months versus eating at the French restaurant in one month followed by eating at home in two months. The majority now wanted the French dinner in two months. For both groups, dinner at home was the most likely alternative to the French dinner, but it was only when the two dinners were expressed as a sequence that the preference for improvement became a basis for decision.

Analyzing how people frame or bracket choices may help illuminate the issue of whether a preference for improvement merely reflects the combined effect of other motives, such as reference dependence or anticipatory utility, or whether it is something unique. Viewed from an integrated decision-making perspective, the preference for improvement seems derivative of these other concepts, because it is unclear why one would value improvement for its own sake. But when viewed from a choice-bracketing perspective, it seems plausible that a person would adopt this choice heuristic for evaluating sequences. Specifically, a preference-for-improvement choice heuristic may have originated from considerations of reference dependence or anticipatory utility, but a person using this choice heuristic may come to feel that improvement for its own sake has value.[25]

Loewenstein and Prelec (1993) develop a choice-heuristic model for how people evaluate choices over sequences. They assume that people consider a sequence's discounted utility, its degree of improvement, and its degree of spread. The key ingredients of the model are "gestalt" definitions for improvement and spread. In other words, they develop a formal measure of the degree of improvement and the degree of spread for any sequence. They show that their model can explain a wide range of sequence anomalies, including observed violations of independence, and that it predicts preferences between sequences much better than other models that incorporate similar numbers of free parameters (even a model with an entirely flexible time-discount function).

MULTIPLE-SELF MODELS

An influential school of theorists has proposed models that view intertemporal choice as the outcome of a conflict between multiple selves. Most multiple-self models postulate myopic selves who are in conflict with more farsighted ones, and often draw analogies between intertemporal choice and a variety of different models of interpersonal strategic interactions. Some models (for example, Ainslie and Haslam 1992; Schelling 1984; Winston 1980) assume that there are two agents, one myopic and one farsighted, who alternately take control of behavior. The main problem with this approach is that it fails to specify why either type of agent emerges when it does. Furthermore, by characterizing the interaction as a battle between the two agents, these models fail to capture an important asymmetry: farsighted selves often attempt to control the behaviors of myopic selves, but never the reverse. For instance, the farsighted self may pour vodka down the drain to prevent tomorrow's self from drinking it, but the myopic self rarely takes steps to ensure that tomorrow's self will have access to the alcohol he or she will then crave.

Responding in part to this problem, Thaler and Shefrin (1981) proposed a "planner-doer" model that draws upon principal-agent theory. In their model, a series of myopic "doers," who care only about their own immediate gratification (and have no affinity for future or past doers), interact with a unitary "planner" who cares equally about the present and future. The model focuses on the strategies employed by the planner to control the behavior of the doers. The model highlights the observation, later discussed at length by Loewenstein (1996), that the farsighted perspective is often much more constant than the myopic perspective.

[25] Thus, to the extent that the preference for improvement reflects a choice heuristic, it should be susceptible to framing or bracketing effects, because what constitutes a sequence is highly subjective, as noted by Loewenstein and Prelec (1993) and by Beebe-Center (1929, p. 67) several decades earlier:

> *What enables one to decide whether a given set of affective experiences does, or does not, constitute a unitary temporal group*? . . . what of series involving experiences of different modalities— . . . visual and auditory experiences, for instance? . . . And what of such complex events as "arising in the morning" or "eating a good meal" or "enjoying a good book?" (emphasis added)

For example, people are often consistent in recognizing the need to maintain a diet. Yet they periodically violate their own desired course of action—often recognizing even at the moment of doing so that they are not behaving in their own self-interest.

Yet a third type of multiple-self model draws connections between intertemporal choice and models of multiperson strategic interactions (Elster 1985). The essential insight that these models capture is that, much like cooperation in a social dilemma, self-control often requires the cooperation of a series of temporally situated selves. When one self "defects" by opting for immediate gratification, the consequence can be a kind of unraveling or "falling off the wagon" when subsequent selves follow the precedent.

Few of these multiple-self models have been expressed formally, and even fewer have been used to derive testable implications that go much beyond the intuitions that inspired them in the first place. However, perhaps it is unfair to criticize the models for these shortcomings. These models are probably best viewed as metaphors intended to highlight specific aspects of intertemporal choice. Specifically, multiple-self models have been used to make sense of the wide range of self-control strategies that people use to regulate their own future behavior. Moreover, these models provided much of the inspiration for more recent formal models of sophisticated hyperbolic discounting (following Laibson 1994, 1997).

TEMPTATION UTILITY

Most models of intertemporal choice—indeed, most models of choice in any framework—assume that options not chosen are irrelevant to a person's well-being. In a recent paper, Gul and Pesendorfer (2001) posit that people have "temptation preferences," wherein they experience disutility from not choosing the option that is most enjoyable now. Their theory implies that a person might be better off if some particularly tempting option were not available, even if he or she doesn't choose that option. As a result, the person may be willing to pay in advance to eliminate that option, or in other words, he or she may have a preference for commitment.

COMBINING INSIGHTS FROM DIFFERENT MODELS

Many behavioral models of intertemporal choice focus on a single modification to the DU model and explore the additional realism produced by that single modification. Yet many empirical phenomena reflect the interaction of multiple phenomena. For instance, a preference for improvement may interact with hyperbolic discounting to produce preferences for U-shaped sequences—for example, for jobs that offer a signing bonus and a salary that increases gradually over time. As discussed by Loewenstein and Prelec (1993), in the short term, the preference-for-improvement motive is swamped by the high discount rates, but as the discount rate falls over time, the preference-for-improvement motive may gain ascendance and cause a net preference for an increasing payment sequence.

As another example, introducing visceral influences into models of hyperbolic discounting may more fully account for the phenomenology of impulsive choices.

Hyperbolic-discounting models predict that people respond especially strongly to immediate costs and benefits, and visceral influences have powerful transient effects on immediate utilities. In combination, the two assumptions could explain a wide range of impulsive choices and other self-control phenomena.

Measuring Time Discounting

The DU model assumes that a person's time preference can be captured by a single discount rate, ρ. In the past three decades there have been many attempts to measure this rate. Some of these estimates are derived from observations of "real-world" behaviors (for example, the choice between electrical appliances that differ in their initial purchase price and long-run operating costs). Others are derived from experimental elicitation procedures (for example, respondents' answers to the question "Which would you prefer: $100 today or $150 one year from today?"). Table 6.1 summarizes the implicit discount rates from all studies that we could locate in which discount rates were either directly reported or easily computed from the reported data.

Figure 6.2 plots the estimated discount factor for each study against the publication date for that study, where the discount factor is $\delta = 1/(1 + \rho)$.[26] This figure reveals three noteworthy observations. First, there is tremendous variability in the estimates (the corresponding implicit annual discount rates range from −6 percent to infinity). Second, in contrast to estimates of physical phenomena such as the speed of light, there is no evidence of methodological progress; the range of estimates is not shrinking over time. Third, high discounting predominates, as most of the data points are well below 1, which represents equal weighting of present and future.

In this section, we provide an overview and critique of this empirical literature with an eye toward understanding these three observations. We then review the procedures used to estimate discount rates. This section reiterates our general theme: To truly understand intertemporal choices, one must recognize the influence of many considerations besides pure time-preference.

Confounding Factors

A wide variety of procedures have been used to estimate discount rates, but most apply the same basic approach. Some actual or reported intertemporal preference is observed, and researchers then compute the discount rate that this preference implies, using a "financial" or net present value (NPV) calculation. For instance, if a person demonstrates indifference between 100 widgets now and 120 widgets in one year, the implicit (annual) discount rate, ρ, would be 20%, because that value would satisfy the equation $100 = (1/(1 + \rho))120$. Similarly, if a person is

[26] In some cases, the estimates are computed from the median respondent. In other cases, the authors reported the mean discount rate.

TABLE 6.1
Empirical Estimates of Discount Rates

Study	*Type*	*Good(s)*	*Real or Hypo?*	*Elicitation Method*	*Time Range*	*Annual Discount Rate(s)*	δ
Maital & Maital (1978)	experimental	money & coupons	hypo	choice	1 year	70%	0.59
Hausman (1979)	field	money	real	choice	undefined	5% to 89%	0.95 to 0.53
Gateley (1980)	field	money	real	choice	undefined	45% to 300%	0.69 to 0.25
Thaler (1981)	experimental	money	hypo	matching	3months to 10 years	7% to 345%	0.93 to 0.22
Ainslie & Haendel (1983)	experimental	money	real	matching	undefined	96000% to ∞	0.00
Houston (1983)	experimental	money	hypo	other	1 year to 20 years	23%	0.81
Loewenstein (1987)	experimental	money & pain	hypo	pricing	immediately to 10 years	−6% to 212%	1.06 to 0.32
Moore and Viscusi (1988)	field	life years	real	choice	undefined	10% to 12%	0.91 to 0.89
Benzion et al. (1989)	experimental	money	hypo	matching	6 months to 4 years	9% to 60%	0.92 to 0.63
Viscusi & Moore (1989)	field	life years	real	choice	undefined	11%	0.90
Moore & Viscusi (1990a)	field	life years	real	choice	undefined	2%	0.98
Moore & Viscusi (1990b)	field	life years	real	choice	undefined	1% to 14%	0.99 to 0.88
Shelley (1993)	experimental	money	hypo	matching	6 months to 4 years	8% to 27%	0.93 to 0.79
Redelmeier & Heller (1993)	experimental	health	hypo	rating	1 day to 10 years	0%	1.00

TABLE 6.1 (*continued*)

Study	*Type*	*Good(s)*	*Real or Hypo?*	*Elicitation Method*	*Time Range*	*Annual Discount Rate(s)*	δ
Cairns (1994)	experimental	money	hypo	choice	5 years to 20 years	14% to 25%	0.88 to 0.80
Shelley (1994)	experimental	money	hypo	rating	6 months to 2 years	4% to 22%	0.96 to 0.82
Chapman & Elstein (1995)	experimental	money & health	hypo	matching	6 months to 12 years	11% to 263%	0.90 to 0.28
Dolan & Gudex (1995)	experimental	health	hypo	other	1 month to 10 years	0%	1.00
Dreyfus and Viscusi (1995)	field	life years	real	choice	undefined	11% to 17%	0.90 to 0.85
Kirby & Marakovic (1995)	experimental	money	real	matching	3 days to 29 days	3678% to ∞	0.03 to 0.00
Chapman (1996)	experimental	money & health	hypo	matching	1 year to 12 years	negative to 300%	1.01 to 0.25
Kirby & Marakovic (1996)	experimental	money	real	choice	6 hours to 70 days	500% to 1500%	0.17 to 0.06
Pender (1996)	experimental	rice	real	choice	7 months to 2 years	26% to 69%	0.79 to 0.59
Wahlund & Gunnarson (1996)	experimental	money	hypo	matching	1 month to 1 year	18% to 158%	0.85 to 0.39
Cairns & Van der Pol (1997)	experimental	money	hypo	matching	2 years to 19 years	13% to 31%	0.88 to 0.76
Green, Myerson, & McFadden (1997)	experimental	money	hypo	choice	3 months to 20 years	6% to 111%	0.94 to 0.47
Johanneson & Johansson (1997)	experimental	life years	hypo	pricing	6 years to 57 years	0% to 3%	0.97

Kirby (1997)	experimental	money	real	pricing	1 day to 1 month	159% to 5747%	0.39 to 0.02
Madden et al. (1997)	experimental	money & heroin	hypo	choice	1 week to 25 years	8% to ∞	0.93 to 0.00
Chapman & Winquist (1998)	experimental	money	hypo	matching	3 months	426% to 2189%	0.19 to 0.04
Holden, Shiferaw, & Wik (1998)	experimental	money & corn	real	matching	1 year	28% to 147%	0.78 to 0.40
Cairns & Van der Pol (1999)	experimental	health	hypo	matching	4 years to 16 years	6%	0.94
Chapman, Nelson, & Hier (1999)	experimental	money & health	hypo	choice	1 month to 6 months	13% to 19000%	0.88 to 0.01
Coller & Williams (1999)	experimental	money	real	choice	1 month to 3 months	15% to 25%	0.87 to 0.80
Kirby, Petry, & Bickel (1999)	experimental	money	real	choice	7 days to 186 days	50% to 55700%	0.67 to 0.00
Van Der Pol & Cairns (1999)	experimental	health	hypo	choice	5 years to 13 years	7%	0.93
Chesson & Viscusi (2000)	experimental	money	hypo	matching	1 year to 25 years	11%	0.90
Ganiats et al. (2000)	experimental	health	hypo	choice	6 months to 20 years	negative to 116%	1.01 to 0.46
Hesketh (2000)	experimental	money	hypo	choice	6 months to 4 years	4% to 36%	0.96 to 0.74
Van Der Pol & Cairns (2001)	experimental	health	hypo	choice	2 years to 15 years	6% to 19%	0.94 to 0.92
Warner & Pleeter (2001)	field	money	real	choice	immediately to 22 years	0% to 71%	0 to 0.58
Harrison, Lau, & Williams (2002)	experimental	money	real	choice	1 month to 37 months	28%	0.78

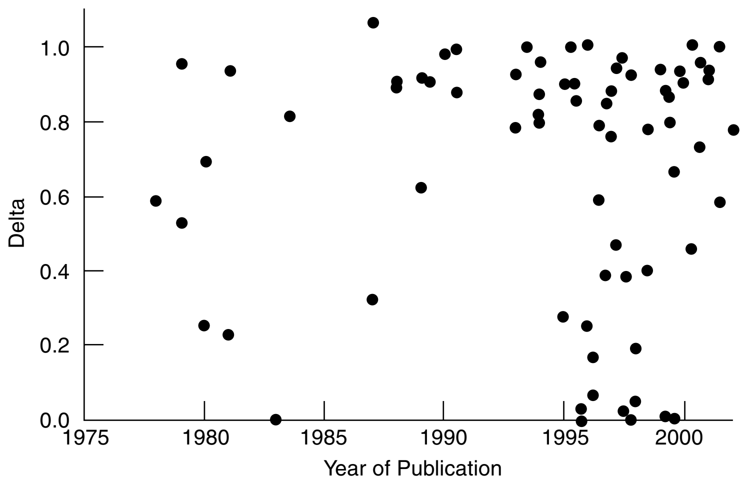


FIGURE 6.2 Discount factor by year of study publication.

indifferent between an inefficient low-cost appliance and a more efficient one that costs $100 extra but saves $20 a year in electricity over the next 10 years, the implicit discount rate, ρ, would equal 15.1%, because that value would satisfy the equation $100 = \Sigma_{t=1}^{10} (1/(1+\rho))^t 20$.

Although this is an extremely widespread approach for measuring discount rates, it relies on a variety of additional (and usually implicit) assumptions, and is subject to several confounding factors.

CONSUMPTION REALLOCATION

The foregoing calculation assumes a sort of "isolation" in decision making. Specifically, it treats the objects of intertemporal choice as discrete, unitary, dated events; it assumes that people entirely "consume" the reward (or penalty) at the moment it is received, as if it were an instantaneous burst of utility. Furthermore, it assumes that people don't shift consumption around over time in anticipation of the receipt of the future reward or penalty. These assumptions are rarely exactly correct, and may sometimes be bad approximations. Choosing between $50 today versus $100 next year, or choosing between 50 pounds of corn today versus 100 pounds next year, are not the same as choosing between 50 utils today and 100 utils on the same day next year, as the calculations imply. Rather, they are more complex choices between the various streams of consumption that those two dated rewards make possible.

INTERTEMPORAL ARBITRAGE

In theory, choices between tradable rewards, such as money, should not reveal anything about time preferences. As Fuchs (1982) and others have noted, if capital markets operate effectively (if monetary amounts at different times can be costlessly exchanged at a specified interest rate), choices between dated monetary outcomes can be reduced to merely selecting the reward with the greatest net present value (using the market interest rate).[27] To illustrate, suppose a person prefers $100 now to $200 ten years from now. While this preference *could* be explained by imputing a discount rate on future utility, the person might be choosing the smaller immediate amount because he or she believes that through proper investment the person can turn it into more than $200 in ten years, and thus enjoy more than $200 worth of consumption *at that future time*. The presence of capital markets should cause imputed discount rates to converge on the market interest rate.

Studies that impute discount rates from choices among tradable rewards assume that respondents ignore opportunities for intertemporal arbitrage, either because they are unaware of capital markets or unable to exploit them.[28] The latter assumption may sometimes be correct. For instance, in field studies of electrical-appliance purchases, some subjects may have faced borrowing constraints that prevented them from purchasing the more expensive energy-efficient appliances. More typically, however, imperfect capital markets cannot explain choices; they cannot explain why a person who holds several thousand dollars in a bank account earning 4 percent interest should prefer $100 today over $150 in one year. Because imputed discount rates in fact do not converge on the prevailing market interest rates, but instead are much higher, many respondents apparently are neglecting capital markets and basing their choices on some other consideration, such as time preference or the uncertainty associated with delay.

CONCAVE UTILITY

The standard approach to estimating discount rates assumes that the utility function is linear in the magnitude of the choice objects (for example, amounts of money, pounds of corn, duration of some health state). If, instead, the utility function for the good in question is concave, estimates of time preference will be biased upward. For example, indifference between $100 this year and $200 next year implies a *dollar* discount rate of 100%. If the utility of acquiring $200 is less than twice the utility of acquiring $100, however, the *utility* discount rate will be less than 100%. This confound is rarely discussed, perhaps because utility is assumed to be approximately linear over the small amounts of money commonly

[27] Meyer (1976, p. 426) expresses this point: "if we can lend and borrow at the same rate . . . , then we can simply show that, regardless of the fundamental orderings on the *c*'s [consumption streams], the induced ordering on the *x*'s [sequences of monetary flows] is given by simple discounting at this given rate. . . . We could say that the market assumes command and the market rate prevails for monetary flows."

[28] Arguments about violations of the discounted utility model assume, as Pender (1996, pp. 282–83) notes, that the results of discount rate experiments reveal something about intertemporal preferences directly. However, if agents are optimizing an intertemporal utility function, their opportunities for intertemporal arbitrage are also important in determining how they respond to such experiments . . . when tradable rewards are offered, one must either abandon the assumption that respondents in experimental studies are optimizing, or make some assumptions (either implicit or explicit) about the nature of credit markets. The implicit assumption in some of the previous studies of discount rates appears to be that there are no possibilities for intertemporal arbitrage.

used in time-preference studies. The overwhelming evidence for reference-dependent utility suggests, however, that this assumption may be invalid—that people may not be integrating the stated amounts with their current and future wealth, and therefore that curvature in the utility function may be substantial even for these small amounts (see Bateman et al. 1997; Harless and Camerer 1994; Kahneman and Tversky 1979; Rabin 2000; Rabin and Thaler 2001; Tversky and Kahneman 1991).

Three techniques could be used to avoid this confound. First, one could request direct utility judgments (for example, attractiveness ratings) of the same consequence at two different times. Then, the ratio of the attractiveness rating of the distant outcome to the proximate outcome would directly reveal the implicit discount factor. Second, to the extent that utility is linear in probability, one can use choices or judgment tasks involving different probabilities of the same consequence at different times (Roth and Murnighan 1982). Evidence that probability is weighted nonlinearly (see, for example, Starmer 2000) would, of course, cast doubt on this approach. Third, one can separately elicit the utility function for the good in question, and then use that function to transform outcome amounts into utility amounts, from which utility discount rates could be computed. To our knowledge, Chapman (1996) conducted the only study that attempted to do this. She found that *utility* discount rates were substantially lower than the *dollar* discount rates, because utility was strongly concave over the monetary amounts subjects used in the intertemporal choice tasks.[29]

UNCERTAINTY

In experimental studies, subjects are typically instructed to assume that delayed rewards will be delivered with certainty. Whether subjects do (or can) accept this assumption is unclear, because delay is ordinarily—and perhaps unavoidably—associated with uncertainty. A similar problem arises for field studies, in which it is typically assumed that subjects believe that future rewards, such as energy savings, will materialize. Due to this subjective (or *epistemic*) uncertainty associated with delay, it is difficult to determine to what extent the magnitude of imputed discount rates (or the shape of the discount function) is governed by time preference per se, versus the diminution in subjective probability associated with delay.[30]

Empirical evidence suggests that introducing objective (or *aleatory*) uncertainty to both current and future rewards can dramatically affect estimated discount rates. For instance, Keren and Roelofsma (1995) asked one group of respondents

[29] Chapman also found that magnitude effects were much smaller after correcting for utility function curvature. This result supports Loewenstein and Prelec's (1992) explanation of magnitude effects as resulting from utility function curvature (see section on reference-point models herein).

[30] There may be complicated interactions between risk and delay, because uncertainty about future receipt complicates and impedes the planning of one's future consumption stream (Spence and Zeckhauser 1972). For example, a 90% chance to win $10,000,000 in 15 years is worth much less than a guarantee to receive $9,000,000 at that time, because, to the extent that one cannot insure against the residual uncertainty, there is a limit to how much one can adjust one's consumption level during those 15 years.

to choose between 100 florins (a Netherlands unit of currency) immediately and 110 florins in one month, and another group to choose between a 50% chance of 100 florins immediately and a 50% chance of 110 florins in one month. While 82% preferred the smaller immediate reward when both rewards were certain, only 39% preferred the smaller immediate reward when both rewards were uncertain.[31] Also, Albrecht and Weber (1996) found that the present value of a future lottery (for example, a 50% chance of receiving 250 deutsche marks) tended to exceed the present value of its certainty equivalent.

INFLATION

The standard approach assumes that, for instance, $100 now and $100 in 5 years generate the same level of utility at the times they are received. However, inflation provides a reason to devalue future monetary outcomes, because in the presence of inflation, $100 worth of consumption now is more valuable than $100 worth of consumption in 5 years. This confound creates an upward bias in estimates of the discount rate, and this bias will be more or less pronounced depending on subjects' experiences with and expectations about inflation.

EXPECTATIONS OF CHANGING UTILITY

A reward of $100 now might also generate more utility than the same amount five years hence because a person expects to have a larger baseline consumption level in 5 years (for example, due to increased wealth). As a result, the marginal utility generated by an additional $100 of consumption in 5 years may be less than the marginal utility generated by an additional $100 of consumption now. Like inflation, this confound creates an upward bias in estimates of the discount rate.

HABIT FORMATION, ANTICIPATORY UTILITY, AND VISCERAL INFLUENCES

To the extent that the discount rate is meant to reflect *only* time preference, and not the confluence of *all* factors influencing intertemporal choice, the modifications to the instantaneous utility function discussed in the previous section represent additional biasing factors, because they are typically not accounted for when the discount rate is imputed. For instance, if anticipatory utility motivates one to delay consumption more than one otherwise would, the imputed discount rate will be lower than the true degree of time preference. If a person prefers an increasing consumption profile due to habit formation, the discount rate will be biased downward. Finally, if the prospect of an immediate reward momentarily stimulates visceral factors that temporarily increase the person's valuation of the proximate reward, the discount rate could be biased upward.[32]

[31] This result cannot be explained by a magnitude effect on the expected amounts, because 50% of a reward has a *smaller* expected value, and, according to the magnitude effect, should be discounted more, not less.

[32] Whether visceral factors should be considered a determinant of time preference or a confounding factor in its estimation is unclear. If visceral factors increase the attractiveness of an immediate reward without affecting its experienced enjoyment (if they increase wanting but not liking), they are probably best viewed as a legitimate determinant of time preference. If, however, visceral factors alter the

AN ILLUSTRATIVE EXAMPLE

To illustrate the difficulty of separating time preference per se from these potential confounds, consider a prototypical study by Benzion, Rapoport, and Yagil (1989). In this study, respondents equated immediate sums of money and larger delayed sums (for example, they specified the reward in six months that would be as good as getting $1,000 immediately). In the cover story for the questionnaire, respondents were asked to imagine that they had earned money (amounts ranged from $40 to $5,000), but when they arrived to receive the payment they were told that the "financially solid" public institute is "temporarily short of funds." They were asked to specify a future amount of money (delays ranged from 6 months to 4 years) that would make them indifferent to the amount they had been promised to receive immediately. Surely, the description "financially solid" could scarcely be sufficient to allay uncertainties that the future reward would actually be received (particularly given that the institute was "temporarily" short of funds), and it seems likely that responses included a substantial "risk premium." Moreover, the subjects in this study had "extensive experience with . . . a three-digit inflation rate," and respondents might well have considered inflation when generating their responses. Even if respondents assumed no inflation, the real interest rate during this time was positive, and they might have considered intertemporal arbitrage. Finally, respondents may have considered that their future wealth would be greater and that the later reward would therefore yield less marginal utility. Indeed, the instructions cued respondents to consider this, as they were told that the questions did not have correct answers, and that the answers "might vary from one individual to another depending on his or her present or future financial assets."

Given all of these confounding factors, it is unclear exactly how much of the imputed annual discount rates (which ranged from 9 to 60%) actually reflected time preference. It is possible that the responses in this study (and others) can be entirely explained in terms of these confounds, and that once these confounds are controlled for, no "pure" time preference would remain.

Procedures for Measuring Discount Rates

Having discussed several confounding factors that greatly complicate assigning a discount rate to a particular choice or judgment, we next discuss the methods that have been used to measure discount rates. Broadly, these methods can be divided into two categories: *field studies*, in which discount rates are inferred from economic decisions people make in their lives, and *experimental studies*, in which people are asked to evaluate stylized intertemporal prospects involving real or hypothetical outcomes. The different procedures are each subject to the confounds discussed earlier and, as shall be seen, are also influenced by a variety of other factors that are theoretically irrelevant, but that can greatly affect the imputed discount rate.

---

amount of utility that a contemplated proximate reward actually delivers, they might best be regarded as a confounding factor.

FIELD STUDIES

Some researchers have estimated discount rates by identifying real-world behaviors that involve trade-offs between the near future and more distant future. Early studies of this type examined consumers' choices among different models of electrical appliances, which presented purchasers with a trade-off between the immediate purchase price and the long-term costs of running the appliance (as determined by its energy efficiency). In these studies, the discount rates implied by consumers' choices vastly exceeded market interest rates and differed substantially across product categories. The implicit discount rate was 17 to 20% for air conditioners (Hausman 1979); 102% for gas water heaters, 138% for freezers, 243% for electric water heaters (Ruderman, Levine, and McMahon 1987); and 45 to 300% for refrigerators, depending on assumptions made about the cost of electricity (Gately 1980).[33]

Another set of studies imputes discount rates from wage-risk trade-offs, in which individuals decide whether to accept a riskier job with a higher salary. Such decisions involve a trade-off between quality of life and expected length of life. The more that future utility is discounted, the less important is length of life, making risky but high-paying jobs more attractive. From such trade-offs, Viscusi and Moore (1989) concluded that workers' implicit discount rate with respect to future life years was approximately 11%. Later, using different econometric approaches with the same data set, Moore and Viscusi (1990a) estimated the discount rates to be around 2%, and Moore and Viscusi (1990b) concluded that the discount rate was somewhere between 1 and 14%. Dreyfus and Viscusi (1995) applied a similar approach to auto-safety decisions and estimated discount rates ranging from 11 to 17%.

In the macroeconomics literature, researchers have imputed discount rates by estimating structural models of life-cycle–saving behavior. For instance, Lawrence (1991) used Euler equations to estimate household time preferences across different socioeconomic groups. She estimated the discount rate of median-income households to be between 4 and 13% depending on the specification. Carroll (1997) criticizes Euler equation estimation on the grounds that most households tend to engage mainly in "buffer-stock" saving early in their lives—they save primarily to be prepared for emergencies—and only conduct "retirement" saving later on. Recent papers have estimated rich, calibrated, stochastic models in which households conduct buffer-stock saving early in life and retirement saving later in life. Using this approach, Carroll and Samwick (1997) report

[33] These findings illustrate how people seem to ignore intertemporal arbitrage. As Hausman (1979) noted, it does not make sense for anyone with positive savings to discount future energy savings at rates higher than the market interest rate. One possible explanation for these results is that people are liquidity constrained. Consistent with such an account, Hausman found that the discount rate varied markedly with income—it was 39% for households with under $10,000 of income, but just 8.9% for households earning between $25,000 and $35,000. Conflicting with this finding, however, a study by Houston (1983, p. 245) that presented individuals with a decision of whether to purchase a hypothetical "energy-saving" device, found that income "played no statistically significant role in explaining the level of discount rate."

point estimates for the discount rate ranging from 5 to 14%, and Gourinchas and Parker (2001) report point estimates of 4.0 to 4.5%. Field studies of this type have the advantage of not assuming isolation, because integrated decision making is built into the model. Yet such estimates often depend heavily on the myriad assumptions included in the structural model.[34]

Recently, Warner and Pleeter (2001) analyzed decisions made by U.S. military servicemen. As part of military downsizing, over 60,000 military employees were given the choice between a onetime, lump-sum payment and an annuity payment. The sizes of the payments depended on the employee's current salary and number of years of service—for example, an "E-5" with 9 years of service could choose between $22,283 now versus $3,714 every year for 18 years. In general, the present value of the annuity payment equaled the lump-sum payment for a discount rate of 17.5%. Although the interest rate was only 7% at the time of these decisions, more than half of all military officers and more than 90% of enlisted personnel chose the lump-sum payment.[35] This study is particularly compelling in terms of credibility of reward delivery, magnitude of stakes, and number of subjects.[36]

The benefit of field studies, as compared with experimental studies, is their high *ecological* validity. There is no concern about whether estimated discount rates would apply to real behavior because they are estimated from such behavior. Yet field studies are subject to additional confounds due to the complexity of real-world decisions and the inability to control for some important factors. For example, the high discount rates implied by the widespread use of inefficient electrical appliances might not result from the discounting of future cost savings per se, but from other considerations, including: a lack of information among consumers about the cost savings of the more efficient appliances; a disbelief among consumers that the cost savings will be as great as promised; a lack of expertise in translating available information into economically efficient decisions; or hidden costs of the more efficient appliances, such as reduced convenience or reliability, or, in the case of lightbulbs, because the more efficient bulbs generate less aesthetically pleasing light spectra.[37]

EXPERIMENTAL STUDIES

Given the difficulties of interpreting field data, the most common methodology for eliciting discount rates is to solicit "paper and pencil" responses to the prospect of real and hypothetical rewards and penalties. Four experimental procedures are commonly used: choice tasks, matching tasks, pricing tasks, and ratings tasks.

[34] These macroeconomics studies are not included in the tables and figures, which focus primarily on individual-level choice data.

[35] It should be noted, however, that the guaranteed payments in the annuity program were not indexed for inflation, which averaged 4.2% during the 4 years preceding this choice.

[36] Warner and Pleeter (2001) noted that if everyone had chosen the annuity payment, the present value of all payments would have been $4.2 billion. Given the choices, however, the present value of the government payout was just $2.5 billion. Thus offering the lump-sum alternative saved the federal government $1.7 billion.

[37] For a criticism of the hidden-costs explanation, however, see Koomey and Sanstad (1994) and Howarth and Sanstad (1995).

Choice tasks are the most common experimental method for eliciting discount rates. In a typical choice task, subjects are asked to choose between a smaller, more immediate reward and a larger, more delayed reward. Of course, a single choice between two intertemporal options only reveals an upper or lower bound on the discount rate—for example, if a person prefers one hundred units of something today over one hundred-twenty units a year from today, the choice merely implies a discount rate of *atleast* 20% per year. To identify the discount rate more precisely, researchers often present subjects with a series of choices that vary the delay or the amount of the rewards. Some studies use real rewards, including money, rice, and corn. Other studies use hypothetical rewards, including monetary gains and losses, and more or less satisfying jobs available at different times. (See table 6.1 for a list of the procedures and rewards used in the different studies.)

Like all experimental elicitation procedures, the results from choice tasks can be affected by procedural nuances. A prevalent problem is an anchoring effect: when respondents are asked to make multiple choices between immediate and delayed rewards, the first choice they face often influences subsequent choices. For instance, people would be more prone to choose $120 next year over $100 immediately if they first chose between $100 immediately and $103 next year than if they first chose between $100 immediately and $140 next year. In general, imputed discount rates tend to be biased in the direction of the discount rate that would equate the first pair of options to which they are exposed (see Green et al. 1998). Anchoring effects can be minimized by using titration procedures that expose respondents to a series of opposing anchors—for example, $100 today or $101 in one year? $100 today or $10,000 in one year? $100 today or $105 in one year? and so on. Since titration procedures typically only offer choices between an immediate reward and a *greater* future reward, however, even these procedures communicate to respondents that they should be discounting, and potentially bias discount rates upward.

Matching tasks are another popular method for eliciting discount rates. In matching tasks, respondents "fill in the blank" to equate two intertemporal options (for example, $100 now = $__ in one year). Matching tasks have been conducted with real and hypothetical monetary outcomes and with hypothetical aversive health conditions (again, see table 6.1 for a list of the procedures and rewards used in different studies). Matching tasks have two advantages over choice tasks. First, because subjects reveal an indifference point, an exact discount rate can be imputed from a single response. Second, because the intertemporal options are not fully specified, there is no anchoring problem and no suggestion of an expected discount rate (or range of discount rates). Thus, unlike choice tasks, matching tasks cannot be accused of simply recovering the expectations of the experimenters that guided the experimental design.

Although matching tasks have some advantages over choice tasks, there are reasons to be suspicious of the responses obtained. First, responses often appear to be governed by the application of some simple rule rather than by time preference. For example, when people are asked to state the amount in n years that equals $100 today, a very common response is $\$100 \times n$. Second, the responses are often very "coarse"—often multiples of 2 or 10 of the immediate reward, suggesting that respondents do not (or cannot) think very carefully about the task. Third, and most important, there are large differences in imputed discount rates among several theoretically equivalent procedures. Two intertemporal options could be equated or matched in one of four ways: respondents could be asked to specify the amount of a delayed reward that would make it as attractive as a given immediate reward (which is the most common technique); the amount of an immediate reward that makes it as attractive as a given delayed reward (Albrecht and Weber 1996); the maximum length of time they would be willing to wait to receive a larger reward in lieu of an immediately available smaller reward (Ainslie and Haendel 1983; Roelofsma 1994); or the latest date at which they would accept a smaller reward in lieu of receiving a larger reward at a specified date that is later still.

While there is no theoretical basis for preferring one of these methods over any other, the small amount of empirical evidence comparing different methods suggests that they yield very different discount rates. Roelofsma (1994) found that implicit discount rates varied tremendously depending on whether respondents matched on amount or time. One group of subjects was asked to indicate how much compensation they would demand to allow a purchased bicycle to be delivered 9 months late. The median response was 250 florins. Another group was asked how long they would be willing to delay delivery of the bicycle in exchange for 250 florins. The mean response was only 3 weeks, implying a discount rate that is 12 times higher. Frederick and Read (2002) found that implicit discount rates were dramatically higher when respondents generated the future reward that would equal a specified current reward than when they generated a current reward that would equal a specified future reward. Specifically, when respondents were asked to state the amount in 30 years that would be as good as getting $100 today, the median response was $10,000 (implying that a future dollar is 1⁄100th as valuable), but when asked to specify the amount today that is as good as getting $100 in thirty years, the median response was $50 (implying that a future dollar is ½ as valuable).

Two other experimental procedures involve rating or pricing temporal prospects. In *rating tasks*, each respondent evaluates an outcome occurring at a particular time by rating its attractiveness or aversiveness. In *pricing tasks*, each respondent specifies a willingness to pay to obtain (or avoid) some real or hypothetical outcome occurring at a particular time, such as a monetary reward, dinner coupons, an electric shock, or an extra year added to the end of one's life. (Once again, see table 6.1 for a list of the procedures and rewards used in the different studies.) Rating and pricing tasks differ from choice and matching tasks in one important respect. Whereas choice and matching tasks call attention to time (because each respondent evaluates two outcomes occurring at two different times), rating and pricing tasks permit time to be manipulated *between subjects* (because a single respondent may evaluate either the immediate or delayed outcome, by itself).

Loewenstein (1988) found that the timing of an outcome is much less important (discount rates are much lower) when respondents evaluate a single outcome at a particular time than when they compare two outcomes occurring at different times, or specify the value of delaying or accelerating an outcome. In one study, for example, two groups of students were asked how much they would pay for a $100 gift certificate at the restaurant of their choice. One group was told that the gift certificate was valid immediately. The other was told it could be used beginning six months from now. There was no significant difference in the valuation of the two certificates *between* the two groups, which implies negligible discounting. Yet when asked how much they would pay (have to be paid) to use it 6 months earlier (later), the timing became important—the delay group was willing to pay $10 to expedite receipt of the delayed certificate, while the immediate group demanded $23 to delay the receipt of a certificate they expected to be able to use immediately.[38]

Another important design choice in experimental studies is whether to use real or hypothetical rewards. The use of real rewards is generally desirable for obvious reasons, but hypothetical rewards actually have some advantages in this domain. In studies involving hypothetical rewards, respondents can be presented with a wide range of reward amounts, including losses and large gains, both of which are generally infeasible in studies involving real outcomes. The disadvantage of hypothetical choice data is the uncertainty about whether people are motivated to, or capable of, accurately predicting what they would do if outcomes were real.

To our knowledge, only two studies have compared discounting between real and hypothetical rewards. Kirby and Marakovic (1995) asked subjects to state the immediate amount that would make them indifferent to some fixed delayed amount (delayed reward sizes were $14.75, $17.25, $21, $24.50, $28.50; delays were 3, 7, 13, 17, 23, and 29 days). One group of subjects answered all 30 permutations for real rewards, and another group of subjects answered all 30 permutations for hypothetical rewards. Discount rates were *lower* for hypothetical rewards.[39] Coller and Williams (1999) asked subjects to choose between $500 payable in 1 month and $500 + $*x* payable in 3 months, where $*x* was varied from $1.67 to $90.94 across 15 different choices. In one condition, all choices were hypothetical; in 5 other conditions, one person was randomly chosen to receive her preferred outcome for 1 of her 15 choices. The raw data suggest again that discount rates were considerably lower in the hypothetical condition, although they suggest that this conclusion is not supported after controlling for censored data, demographic differences, and heteroskedasticity (across demographic differences and across

[38] Rating tasks (and probably pricing tasks as well) are subject to anchoring effects. Shelley and Omer (1996), Stevenson (1992), and others have found that a given delay (for example, 6 months) produces greater time discounting when it is considered alongside shorter delays (for example, 1 month) than when it is considered alongside longer delays (for example, 3 years).

[39] The two results were not strictly comparable, however, because they used a different procedure for the real rewards than for the hypothetical rewards. An auction procedure was used for the real-rewards group only. Subjects were told that whoever, of three subjects, stated the lowest immediate amount would receive the immediate amount, and the other two subjects would receive the delayed amount. Optimal behavior in such a situation involves overbidding. Since this creates a downward bias in discount rates for the real-rewards group, however, it does not explain away the finding that real discount rates were higher than hypothetical discount rates.

treatments).[40] Thus, as of yet there is no clear evidence that hypothetical rewards are discounted differently than real rewards.[41]

What Is Time Preference?

Figure 6.2 reveals spectacular disagreement among dozens of studies that all purport to be measuring time preference. This lack of agreement likely reflects the fact that the various elicitation procedures used to measure time preference consistently fail to isolate time preference, and instead reflect, to varying degrees, a blend of both pure time preference and other theoretically distinct considerations, including: intertemporal arbitrage, when tradeable rewards are used; concave utility; uncertainty that the future reward or penalty will actually obtain; inflation, when nominal monetary amounts are used; expectations of changing utility; and considerations of habit formation, anticipatory utility, and visceral influences.

Figure 6.2 also reveals a predominance of high implicit discount rates—discount rates well above market interest rates. This consistent finding may also be due to the presence of the aforementioned various extra-time-preference considerations, because nearly all of these work to bias imputed discount rates upward—only habit formation and anticipatory-utility bias estimates downward. If these confounding factors were adequately controlled, we suspect that many intertemporal choices or judgments would imply much lower—indeed, possibly even zero—rates of time preference.

Our discussion in this section highlights the conceptual and semantic ambiguity about what the concept of time preference ought to include—about what properly counts as time preference per se and what ought to be called something else (for further discussion see Frederick 1999). We have argued here that many of the reasons for caring when something occurs (for example, uncertainty or utility of anticipation) are not time preference, because they pertain to the expected amount of utility consequences confer, and not to the weight given to the utility of different moments (see figure 6.3 adapted from Frederick 1999). However, it is not obvious where to draw the line between factors that operate through utilities and factors that make up time preference.

Hopefully, economists will eventually achieve a consensus about what is included in, and excluded from, the concept of time preference. Until then, drawing attention to the ambiguity of the concept should improve the quality of discourse

[40] It is hard to understand which control eliminates the differences that are apparent in the raw data. It would seem not to be the demographic differences per se, because the hypothetical condition had a "substantially higher proportion of non-white participants" and "non-whites on average reveal discount rates that are nearly 21 percentage points higher than those revealed by whites" (Coller and Williams 1999, pp. 121, 122).

[41] There has been considerable recent debate outside of the context of intertemporal choice about whether hypothetical choices are representative of decisions with real consequences. The general conclusion from this debate is that the two methods typically yield qualitatively similar results (see Camerer and Hogarth 1999 for a recent review), though systematic differences have been observed in some studies (Cummings, Harrison, and Rutstrom 1995; Kroll, Levy, and Rapoport 1988).

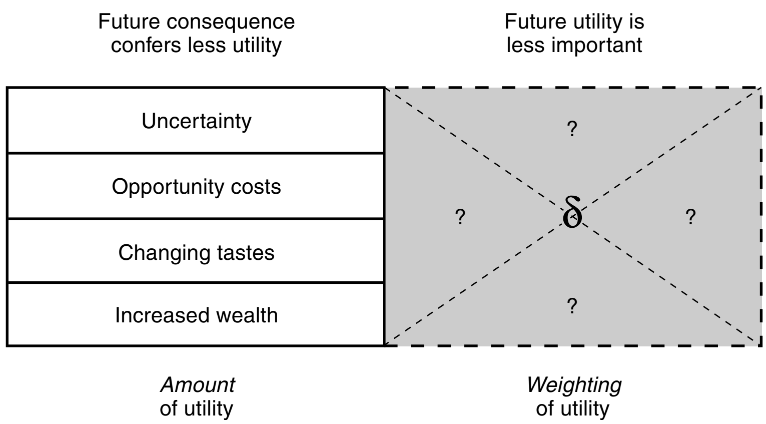


FIGURE 6.3 Amount and weighting of future utility.
Source: Adapted from Fredrick (1999).

by increasing awareness that, in discussions about time preference, different people may be using the same term to refer to significantly different underlying constructs.[42]

UNPACKING TIME PREFERENCE

Early twentieth-century economists' conceptions of intertemporal choice included detailed accounts of disparate underlying psychological motives. With the advent of the DU model in 1937, however, economists eschewed considerations of specific motives, proceeding as if all intertemporal behavior could be explained by the unitary construct of time preference. In this section, we question whether even time preference itself should be regarded as a unitary construct.

Issues of this type are hotly debated in psychology. For example, psychologists debate the usefulness of conceptualizing intelligence in terms of a single unitary

[42] Not only do people use the same term to refer to different concepts (or sets of concepts), they also use different terms to represent the same concept. The welter of terms used in discussions of intertemporal choice include discount factor, discount rate, marginal private rate of discount, social discount rate, utility discount rate, marginal social rate of discount, pure discounting, time preference, subjective rate of time preference, pure time preference, marginal rate of time preference, social rate of time preference, overall time preference, impatience, time bias, temporal orientation, consumption rate of interest, time positivity inclination, and "the pure futurity effect." Broome (1995, pp. 128–29) notes that some of the controversy about discounting results from differences in how the term is used:

> On the face of it . . . typical economists and typical philosophers seem to disagree. But actually I think there is more misunderstanding here than disagreement. . . . When economists and philosophers think of discounting, they typically think of discounting different things. Economists typically discount the sorts of goods that are bought and sold in markets [whereas] philosophers are typically thinking of a more fundamental good, people's *well-being*. . . . It is perfectly consistent to discount commodities and not well-being.

"g" factor. Typically, a posited psychological construct (or "trait") is considered useful only if it satisfies three criteria: it remains relatively constant across time within a particular individual; it predicts behavior across a wide range of situations, and different measures of it correlate highly with one another. The concept of intelligence satisfies these criteria fairly well.[43] First, performance in tests of cognitive ability at early ages correlates highly with performance on such tests at all subsequent ages. Second, cognitive ability (as measured by such tests) predicts a wide range of important life outcomes, such as criminal behavior and income. Third, abilities that we regard as expressions of intelligence correlate strongly with each other. Indeed, when discussing the construction of intelligence tests, Herrnstein and Murray (1994, 3) note, "It turned out to be nearly impossible to devise items that plausibly measured some cognitive skill and were *not* positively correlated with other items that plausibly measured some cognitive skill."

The posited construct of time preference does not fare as well by these criteria. First, no longitudinal studies have been conducted to permit any conclusions about the temporal stability of time preference.[44] Second, correlations between various measures of time preference or between measures of time preference and plausible real-world expressions of it are modest, at best. Chapman and Elstein (1995) and Chapman, Nelson, and Hier (1999) found only weak correlations between discount rates for money and for health, and Chapman and Elstein found almost no correlation between discount rates for losses and for gains. Fuchs (1982) found no correlation between a prototypical measure of time preference (for example, "Would you choose $1,500 now or $4,000 in five years?") and other behaviors that would plausibly be affected by time preference (for example, smoking, credit card debt, seat belt use, and the frequency of exercise and dental checkups). Nor did he find much correlation among any of these reported behaviors (see also Nyhus 1995).[45] Chapman and Coups (1999) found that corporate employees who chose to receive an influenza vaccination did have significantly lower discount rates (as inferred from a matching task with monetary losses), but

[43] Debates remain, however, about whether traditional measures exclude important dimensions, and whether a multidimensional account of intelligence would have even greater explanatory power. Sternberg (1985), for example, argues that intelligence is usefully decomposed into three dimensions: analytical intelligence, which includes the ability to identify problems, compute strategies, and monitor solutions, and is measured well by existing IQ tests; creative intelligence, which reflects the ability to generate problem-solving options, and practical intelligence, which involves the ability to implement problem-solving options.

[44] Although there have been no longitudinal studies of time preference per se, Mischel and his colleagues did find that a child's capacity to delay gratification was significantly correlated with *other* variables assessed decades later, including academic achievement and self-esteem (Ayduk et al. 2000; Mischel, Shoda, and Peake 1988; Shoda, Mischel, and Peake 1990). Of course, this provides evidence for construct validity only to the extent that one views these other variables as expressions of time preference. We also note that while there is little evidence that intertemporal behaviors are stable over long periods, there is some evidence that time preference is not strictly constant over time for all people. Heroin addicts discount both drugs and money more steeply when they are craving heroin than when they are not (Giordano et al. 2001).

[45] A similar lack of *intraindividual* consistency has been observed in risk taking (MacCrimmon and Wehrung 1990).

found no relation between vaccination behavior and hypothetical questions involving health outcomes. Munasinghe and Sicherman (2000) found that smokers tend to invest less in human capital (they have flatter wage profiles), and many others have found that for stylized intertemporal choices among monetary rewards, heroin addicts have higher discount rates (for example, Alvos, Gregson, and Ross 1993; Kirby, Petry, and Bickel 1999; Madden et al. 1997; Murphy and De Wolfe 1986; Petry, Bickel, and Arnett 1998).

Although the evidence in favor of a single construct of time preference is hardly compelling, the low cross-behavior correlations do not necessarily disprove the existence of time preference. Suppose, for example, that someone expresses low discount rates on a conventional elicitation task, yet indicates that she rarely exercises. While it is possible that this inconsistency reflects true heterogeneity in the degree to which she discounts different types of utility, perhaps she rarely exercises *because* she is so busy at work earning money for her future or because she simply cares much more about her future finances than her future cardiovascular condition. Or, perhaps she doesn't believe that exercise improves health. As this example suggests, many factors could work to erode cross-behavior correlations, and thus, such low correlations do not mean that there can be no single unitary time preference underlying all intertemporal choices (the intertemporal analog to the hypothesized construct of "g" in analyses of cognitive performance). Notwithstanding this disclaimer, however, in our view the cumulative evidence raises serious doubts about whether in fact there is such a construct—a stable factor that operates identically on, and applies equally to, all sources of utility.[46]

To understand better the pattern of correlations in implied discount rates across different types of intertemporal behaviors, we may need to unpack time preference itself into more fundamental motives, as illustrated by the segmentation of the delta component of figure 6.3. Loewenstein and his colleagues (2001) have proposed three specific constituent motives, which they labeled *impulsivity* (the degree to which an individual acts in a spontaneous, unplanned fashion), *compulsivity* (the tendency to make plans and stick with them), and *inhibition* (the ability to inhibit the automatic or "knee-jerk" response to the appetites and emotions that trigger impulsive behavior).[47] Preliminary evidence suggests that these subdimensions of time preference can be measured reliably. Moreover, the different subdimensions predict different behaviors in a highly sensible way. For example, repetitive behaviors such as flossing one's teeth, exercising, paying one's bills on

[46] Note that one can also *over*estimate the strength of the relationship between measured time preference and time-related behaviors or between different time-related behaviors if these variables are related to characteristics such as intelligence, social class, or social conformity, that are not adequately measured and controlled for.

[47] Recent research by Baumeister, Heatherton, and Tice (1994) suggests that such "behavioral inhibition" requires an expenditure of mental effort that, like other forms of effort, draws on limited resources—a "pool" of willpower (Loewenstein 2000a). Their research shows that behavioral inhibition in one domain (for example, refraining from eating desirable food) reduces the ability to exert willpower in another domain (for example, completing a taxing mental or physical task).

time, and arriving on time at meetings were all predicted best by the compulsivity subdimension. Viscerally driven behaviors, such as reacting aggressively to someone in a car who honks at you at a red light, were best predicted by impulsivity (positively) and behavioral inhibition (negatively). Money-related behaviors such as saving money, having unpaid credit card balances, or being maxed out on one or more credit cards were best predicted by conventional measures of discount rates (but impulsivity and compulsivity were also highly significant predictors).

Clearly, further research is needed to evaluate whether time preference is best viewed as a unitary construct or a composite of more basic constituent motives. Further efforts hopefully will be informed by recent discoveries of neuroscientists, who have identified regions of the brain whose damage leads to extreme myopia (Damasio 1994) and areas that seem to play an important role in suppressing the behavioral expression of urges (LeDoux 1996). If some behaviors are best predicted by impulsivity, some by compulsivity, some by behavioral inhibition, and so on, it may be worth the effort to measure preferences at this level and to develop models that treat these components separately. Of course, such multidimensional perspectives will inevitably be more difficult to operationalize than formulations like the DU model, which represent time preference as a unidimensional construct.

Conclusions

The DU model, which continues to be widely used by economists, has little empirical support. Even its developers—Samuelson who originally proposed the model, and Koopmans, who provided the first axiomatic derivation—had concerns about its descriptive realism, and it was never empirically validated as the appropriate model for intertemporal choice. Indeed, virtually every core and ancillary assumption of the DU model has been called into question by empirical evidence collected in the past two decades. The insights from this empirical research have spawned new theories of intertemporal choice that revive many of the psychological considerations discussed by early students of intertemporal choice—considerations that were effectively dismissed with the introduction of the DU model. Additionally, some of the most recent theories show that intertemporal behaviors may be dramatically influenced by people's level of understanding of how their preferences change—by their "metaknowledge" about their preferences (see for example, O'Donoghue and Rabin 1999b; Loewenstein, O'Donoghue, and Rabin 2000).

While the DU model assumes that intertemporal preferences can be characterized by a single discount rate the large empirical literature devoted to measuring discount rates has failed to establish any stable estimate. There is extraordinary variation across studies, and sometimes even within studies. This failure is partly due to variations in the degree to which the studies take account of factors that confound the computation of discount rates (for example, uncertainty about the delivery of future outcomes or nonlinearity in the utility function). But the spectacular cross-study differences in discount rates also reflect the diversity of considerations that are relevant in inter-temporal choices and that legitimately affect different types of intertemporal choices differently. Thus there is no reason to expect that discount rates *should* be consistent across different choices.

The idea that intertemporal choices reflect an interplay of disparate and often competing psychological motives was commonplace in the writings of early twentieth-century economists. We believe that this approach should be resurrected. Reintroducing the multiple-motives approach to intertemporal choice will help us to better understand and better explain the intertemporal choices we observe in the real world. For instance, it permits more scope for understanding individual differences (for example, why one person is a spendthrift while his neighbor is a miser, or why one person does drugs while her brother does not), because people may differ in the degree to which they experience anticipatory utility or are influenced by visceral factors.

The multiple-motive approach may be even more important for understanding *intra*individual differences. When one looks at the behavior of a single individual across different domains, there is often a wide range of apparent attitudes toward the future. Someone may smoke heavily, but carefully study the returns of various retirement packages. Another may squirrel money away while at the same time giving little thought to electrical efficiency when purchasing an air conditioner. Someone else may devote two decades of his life to establishing a career, and then jeopardize this long-term investment for some highly transient pleasure. Since the DU model assumes a unitary discount rate that applies to all acts of consumption, such intraindividual heterogeneities pose a theoretical challenge. The multiple-motive approach, by contrast, allows us to readily interpret such differences in terms of more narrow, more legitimate, and more stable constructs—for example, the degree to which people are skeptical of promises, experience anticipatory utility, are influenced by visceral factors, or are able to correctly predict their future utility.

The multiple-motive approach may sound excessively open-ended. We have described a variety of considerations that researchers could potentially incorporate into their analyses. Including every consideration would be far too complicated, while picking and choosing which considerations to incorporate may leave one open to charges of being ad hoc. How, then, should economists proceed?

We believe that economists should proceed as they typically do. Economics has always been both an art and a science. Economists are forced to intuit, to the best of their abilities, which considerations are likely to be important in a particular domain and which are likely to be largely irrelevant. When economists model labor supply, for instance, they typically do so with a utility function that incorporates consumption and leisure, but when they model investment decisions, they typically assume that preferences are defined over wealth. Similarly, a researcher investigating charitable giving might use a utility function that incorporates altruism but not risk aversion or time preference, whereas someone studying investor behavior is unlikely to use a utility function that incorporates altruism. For each domain, economists choose the utility function that is best able to incorporate the essential considerations for that domain, and then evaluate whether the inclusion of specific considerations improves the predictive or explanatory power of a model. The same approach can be applied to multiple-motive models of intertemporal choice. For drug addiction, for example, habit formation, visceral factors, and hyperbolic discounting seem likely to play a prominent role. For extended experiences, such as health states, careers, and long vacations, the preference for improvement is likely to come into play. For brief, vivid experiences, such as weddings or criminal sanctions, utility from anticipation may be an important determinant of behavior.

In sum, we believe that economists' understanding of intertemporal choices will progress most rapidly by continuing to import insights from psychology, by relinquishing the assumption that the key to understanding intertemporal choices is finding the right discount rate (or even the right discount function), and by readopting the view that intertemporal choices reflect many distinct considerations and often involve the interplay of several competing motives. Since different motives may be evoked to different degrees by different situations (and by different descriptions of the *same* situation), developing descriptively adequate models of intertemporal choice will not be easy; but we hope this discussion will help.

References

Abel, Andrew. 1990. "Asset Prices Under Habit Formation and Catching Up with the Joneses." *American Economic Review*, 80: 38–42.

Ainslie, George. 1975. "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin*, 82(4): 463–96.

Ainslie, George, and Varda Haendel. 1983. "The Motives of the Will." In *Etiologic Aspects of Alcohol and Drug Abuse*, edited by E. Gottheil, K. Durley, T. Skodola, and H. Waxman. Springfield, Ill.: Charles C. Thomas.

Ainslie, George, and Nick Haslam. 1992. "Hyperbolic Discounting." In *Choice Over Time*, edited by George Loewenstein and Jon Elster. New York: Russell Sage.

Ainslie, George, and Richard J. Herrnstein. 1981. "Preference Reversal and Delayed Reinforcement." *Animal Learning Behavior*, 9(4): 476–82.

Akerlof, George A. 1991. "Procrastination and Obedience." *American Economic Review*, 81(2): 1–19.

Albrecht, Martin, and Martin Weber. 1995. "Hyperbolic Discounting Models in Prescriptive Theory of Intertemporal Choice." *Zeitschrift Fur Wirtschafts-U Sozialwissenschaften*, 115(S): 535–68.

———. 1996. "The Resolution of Uncertainty: An Experimental Study." *Journal of Theoretical Economics*, 152(4): 593–607.

Alvos, Leanne, R. A. Gregson, and Michael W. Ross. 1993. "Future Time Perspective in Current and Previous Injecting Drug Users." *Drug and Alcohol Dependency*, 31: 193–97.

Angeletos, George-Marios, David Laibson, Andrea Repetto, Jeremy Tobacman, and Stephen Weinberg. 2001. "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation." *Journal of Economic Perspectives*, 15(3): 47–68.

Ariely, Daniel, and Klaus Wertenbroch. 2002. "Procrastination, Deadlines, and Performance: Using Precommitment to Regulate One's Behavior." *Psychological Science*, 13(3): 219–24.

Arrow, Kenneth J. 1983. "The Trade-Off Between Growth and Equity." In *Social Choice and Justice: Collected Papers of Kenneth J. Arrow*, edited by Kenneth J. Arrow. Cambridge, MA: Belknap Press.

Ayduk, Ozlem, Rodolfo Mendoza-Denton, Walter Mischel, G. Downey, Philip K. Peake, and Monica Rodriguez. 2000. "Regulating the Interpersonal Self: Strategic Self-Regulation for Coping with Rejection Sensitivity." *Journal of Personality and Social Psychology*, 79(5): 776–92.

Bateman, Ian, Alistair Munro, Bruce Rhodes, Chris Starmer, and Robert Sugden. 1997. "A Test of the Theory of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 112(2): 479–505.

Baumeister, Roy F., Todd F. Heatherton, and Diane M. Tice. 1994. *Losing Control: How and Why People Fail at Self-Regulation*. San Diego: Academic Press.

Becker, Gary, and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy*, 96(4): 675–701.

Beebe-Center, John G. 1929. "The Law of Affective Equilibrium." *American Journal of Psychology*, 41: 54–69.

Benabou, Roland, and Jean Tirole. 2000. "Self-Confidence: Intrapersonal Strategies." Discussion paper 209, Princeton University.

Benartzi, Shlomo, and Richard H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics*, 110(1): 73–92.

Benzion, Uri, Amnon Rapoport, and Joseph Yagil. 1989. "Discount Rates Inferred from Decisions: An Experimental Study." *Management Science*, 35: 270–84.

Bernheim, Douglas, and Antonio Rangel. 2001. "Addiction, Conditioning, and the Visceral Brain." Unpublished paper, Stanford University, Palo-Alto, CA.

Böhm-Bawerk, Eugen von. 1970 [1889]. *Capital and Interest*. South Holland: Libertarian Press.

Boldrin, Michele, Lawrence Christiano, and Jonas Fisher. 2001. "Habit Persistence, Asset Returns, and the Business Cycle." *American Economic Review*, 9l: 149–66.

Bowman, David, Deborah Minehart, and Matthew Rabin. 1999. "Loss Aversion in a Consumption-Savings Model." *Journal of Economic Behavior and Organization*, 38(2): 155–78.

Broome, John. 1995. "Discounting the Future." *Philosophy & Public Affairs*, 20: 128–56.

Cairns, John A. 1992. "Discounting and Health Benefits." *Health Economics*, 1: 76–79.

———. 1994. "Valuing Future Benefits." *Health Economics*, 3: 221–29.

Cairns, John A., and Marjon M. van der Pol. 1997. "Constant and Decreasing Timing Aversion for Saving Lives." *Social Science and Medicine*, 45(11): 1653–59.

———. 1999. "Do People Value Their Own Future Health Differently than Others' Future Health?" *Medical Decision Making*, 19(4): 466–72.

Camerer, Colin F., and Robin M. Hogarth. 1999. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor Production Framework." *Journal of Risk Uncertainty*, 19: 7–42.

Campbell, John, and John Cochrane. 1999. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy*, 107: 205–51.

Caplin, Andrew, and John Leahy. 2001. "Psychological Expected Utility Theory and Anticipatory Feelings." *Quarterly Journal of Economics*, l66: 55–79.

Carrillo, Juan D. 1999. “Self-Control, Moderate Consumption, and Craving.” CEPR discussion paper 2017.

Carrillo, Juan D., and Thomas Mariotti. 2000. “Strategic Ignorance as a Self-Disciplining Device.” *Review of Economic Studies*, 67(3): 529–44.

Carroll, Christopher. 1997. “Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis.” *Quarterly Journal of Economics*, 112: 1–55.

Carroll, Christopher, Jody Overland, and David Weil. 2000. “Saving and Growth with Habit Formation.” *American Economic Review*, 90: 341–55.

Carroll, Christopher, and Andrew Samwick. 1997. “The Nature of Precautionary Wealth.” *Journal of Monetary Economics*, 40: 41–71.

Chakravarty, S. 1962. “The Existence of an Optimum Savings Program.” *Econometrica*, 30(1): 178–87.

Chapman, Gretchen B. 1996. “Temporal Discounting and Utility for Health and Money.” *Journal of Experimental Psychology: Learning, Memory, Cognition*, 22(3): 771–91.

———. 2000. “Preferences for Improving and Declining Sequences of Health Outcomes.” *Journal of Behavioral Decision Making*, 13: 203–18.

Chapman, Gretchen B., and Elliot J. Coups. 1999. “Time Preferences and Preventive Health Behavior: Acceptance of the Influenza Vaccine.” *Medical Decision Making*, 19(3): 307–14.

Chapman, Gretchen B., and Arthur S. Elstein. 1995. “Valuing the Future: Temporal Discounting of Health and Money.” *Medical Decision Making*, 15(4): 373–86.

Chapman, Gretchen, Richard Nelson, and Daniel B. Hier. 1999. “Familarity and Time Preferences: Decision Making About Treatments for Migraine Headaches and Crohn’s Disease.” *Journal of Experimental Psychology: Applied*, 5(1): 17–34.

Chapman, Gretchen B., and Jennifer R. Winquist. 1998. “The Magnitude Effect: Temporal Discount Rates and Restaurant Tips.” *Psychonomic Bulletin and Review*, 5(1): 119–23.

Chesson, Harrell, and W. Kip Viscusi. 2000. “The Heterogeneity of Time-Risk Trade-offs.” *Journal of Behavioral Decision Making*, 13: 251–58.

Coller, Maribeth, and Melonie B. Williams. 1999. “Eliciting Individual Discount Rates.” *Experimental Economy*, 2: 107–27.

Constantinides, George M. 1990. “Habit Formation: A Resolution of the Equity Premium Puzzle.” *Journal of Political Economy*, 98(3): 519–43.

Cummings, Ronald G., Glenn W. Harrison, and E. Elisabet Rutstrom. 1995. “Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible?” *American Economic Review*, 85: 260–66.

Damasio, Antonio R. 1994. *Descartes’ Error: Emotion, Reason, and the Human Brain*. New York: G. P. Putnam.

Dolan, Paul, and Claire Gudex. 1995. “Time Preference, Duration and Health State Valuations.” *Health Economics*, 4: 289–99.

Dreyfus, Mark K., and W. Kip Viscusi. 1995. “Rates of Time Preference and Consumer Valuations of Automobile Safety and Fuel Efficiency.” *Journal of Law and Economics*, 38(1): 79–105.

Duesenberry, James. 1952. *Income, Saving, and the Theory of Consumer Behavior*. Cambridge: Harvard University Press.

Elster, Jon, 1979. *Ulysses and the Sirens: Studies in Rationality and Irrationality*. Cambridge: Cambridge University Press.

———. 1985. “Weakness of Will and the Free-Rider Problem.” *Economics and Philosophy*, 1: 231–65.

Fischer, Carolyn. 1999. "Read This Paper Even Later: Procrastination with Time-Inconsistent Preferences." Discussion paper 99-20, Resources for the Future.

Fishburn, Peter C. 1970. *Utility Theory and Decision Making*. New York: Wiley.

Fishburn, Peter C., and Ariel Rubinstein. 1982. "Time Preference." *International Economic Review*, 23(2): 677–94.

Fisher, Irving. 1930. *The Theory of Interest*. New York: Macmillan.

Frank, Robert. 1993. "Wages, Seniority, and the Demand for Rising Consumption Profiles." *Journal of Economic Behavior and Organization*, 21: 251–76.

Frederick, Shane. 1999. "Discounting, Time Preference, and Identity." Ph.D. diss., Carnegie Mellon University.

Frederick, Shane, and George Loewenstein. 2002. "The Psychology of Sequence Preferences." Working paper, Sloan School, MIT, MA.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature*, 40(June): 351–401.

Frederick, Shane, and Daniel Read. 2002. "The Empirical and Normative Status of Hyperbolic Discounting and Other DU Anomalies." Working paper, MIT and London School of Economics, Cambridge and London.

Fuchs, Victor. 1982. "Time Preferences and Health: An Exploratory Study." In *Economic Aspects of Health*, edited by Victor Fuchs. Chicago: University of Chicago Press.

Fuhrer, Jeffrey. 2000. "Habit Formation in Consumption and Its Implications for Monetary-Policy Models." *American Economic Review*, 90: 367–90.

Ganiats, Theodore G., Richard T. Carson, Robert M. Hamm, Scott B. Cantor, Walton Sumner, Stephen J. Spann, Michael Hagen, and Christopher Miller. 2000. "Health Status and Preferences: Population-Based Time Preferences for Future Health Outcome." *Medical Decision Making*, 20(3): 263–70.

Gately, Dermot. 1980. "Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables: Comment." *Bell Journal of Economics*, 11: 373–74.

Giordano, Louis A., Warren Bickel, George Loewenstein, Eric Jacobs, Lisa Marsch, and Gary J. Badger. 2001. "Opioid Deprivation Affects How Opioid-Dependent Outpatients Discount the Value of Delayed Heroin and Money." Working paper, Psychiatry Department Substance Abuse Treatment Center, University of Vermont, Burlington.

Goldman, Steven M. 1980. "Consistent Plans." *Review Economic Studies*, 47(3): 533–37.

Gourinchas, Pierre-Olivier, and Jonathan Parker. 2001. "The Empirical Importance of Precautionary Saving." *American Economic Review*, 91(2): 406–12.

Green, Donald, Karen Jacowitz, Daniel Kahneman, and Daniel Mcfadden. 1998. "Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods." *Resource Energy Economics*, 20: 85–116.

Green, Leonard, E. B. Fischer, Jr., Steven Perlow, and Lisa Sherman. 1981. "Preference Reversal and Self Control: Choice as a Function of Reward Amount and Delay." *Behavioral Analysis Letters*, 1(1): 43–51.

Green, Leonard, Nathanael Fristoe, and Joel Myerson. 1994. "Temporal Discounting and Preference Reversals in Choice Between Delayed Outcomes." *Psychonomic Bulletin and Review*, 1(3): 383–89.

Green, Leonard, Astrid Fry, and Joel Myerson. 1994. "Discounting of Delayed Rewards: A Life-Span Comparison." *Psychological Science*, 5(1): 33–36.

Green, Leonard, Joel Myerson, and Edward McFadden. 1997. "Rate of Temporal Discounting Decreases with Amount of Reward." *Memory & Cognition*, 25(5): 715–23.

Gruber, Jonathan, and Botond Koszegi. 2000. "Is Addiction 'Rational'? Theory and Evidence." NBER working paper 7507.

Gul, Faruk, and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control." *Econometrica*, 69: 1403–35.

Harless, David W., and Colin F. Camerer. 1994. "The Predictive Utility of Generalized Expected Utility Theories." *Econometrica*, 62(6): 1251–89.

Harrison, Glenn W., Morten I. Lau, and Melonie B. Williams. 2002. "Estimating Individual Discount Rates in Denmark." *Quarterly Journal of Economics*, 116(4): 1261–1303.

Hausman, Jerry. 1979. "Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables." *Bell Journal of Economics*, 10(1): 33–54.

Hermalin, Benjamin, and Alice Isen. 2000. "The Effect of Affect on Economic and Strategic Decision Making." Mimeo, University of California, Berkeley, and Cornell University.

Herrnstein, Richard. 1981. "Self-Control as Response Strength." In *Quantification of Steady-State Operant Behavior*, edited by Christopher M. Bradshaw, Elmer Szabadi, and C. F. Lowe. North Holland: Elsevier.

Herrnstein, Richard J., George F. Loewenstein, Drazen Prelec, and William Vaughan. 1993. "Utility Maximization and Melioration: Internalities in Individual Choice." *Journal of Behavioral Decision Making*, 6(3): 149–85.

Herrnstein, Richard J., and Charles Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hesketh, Beryl. 2000. "Time Perspective in Career-Related Choices: Applications of Time-Discounting Principles." *Journal of Vocational Behavior*, 57: 62–84.

Hirshleifer, Jack. 1970. *Investment, Interest, and Capital*. Englewood Cliffs, N.J.: Prentice-Hall.

Holcomb, James H., and Paul S. Nelson. 1992. "Another Experimental Look at Individual Time Preference." *Rationality Society*, 4(2): 199–220.

Holden, Stein T., Bekele Shiferaw, and Mette Wik. 1998. "Poverty, Market Imperfections and Time Preferences of Relevance for Environmental Policy?" *Environmental Development Economics*, 3: 105–30.

Houston, Douglas A. 1983. "Implicit Discount Rates and the Purchase of Untried, Energy-Saving Durable Goods." *Journal of Consumer Resources*, 10: 236–46.

Howarth, Richard B., and Alan H. Sanstad. 1995. "Discount Rates and Energy Efficiency." *Contemporary Economic Policy*, 13(3): 101–109.

Hsee, Christopher K., Robert P. Abelson, and Peter Salovey. 1991. "The Relative Weighting of Position and Velocity in Satisfaction." *Psychological Science* 2(4): 263–66.

Jermann, Urban. 1998. "Asset Pricing in Production Economies." *Journal of Monetary Economics*, 41: 257–75.

Jevons, Herbert S. 1905. *Essays on Economics*. London: Macmillan.

Jevons, William S. 1888. *The Theory of Political Economy*. London: Macmillan.

Johannesson, Magnus, and Per-Olov Johansson. 1997. "Quality of Life and the WTP for an Increased Life Expectancy at an Advanced Age." *Journal of Public Economics*, 65: 219–28.

Kahneman, Daniel. 1994. "New Challenges to the Rationality Assumption." *Journal of Institutional and Theoretical Economics*, 150: 18–36.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, 47: 263–92.

Kahneman, Daniel, Peter Wakker, and Rakesh Sarin. 1997. "Back to Bentham? Explorations of Experienced Utility." *Quarterly Journal of Economics*, 112: 375–405.

Keren, Gideon, and Peter Roelofsma. 1995. “Immediacy and Certainty in Intertemporal Choice.” *Organizational Behavior and Human Decision Processes*, 63(3): 287–97.

Kirby, Kris N. 1997. “Bidding on the Future: Evidence Against Normative Discounting of Delayed Rewards.” *Journal of Experimental Psychology: General*, 126: 54–70.

Kirby, Kris N., and Richard J. Herrnstein. 1995. “Preference Reversals Due to Myopic Discounting of Delayed Reward.” *Psychological Science*, 6(2): 83–89.

Kirby, Kris N., and Nino N. Marakovic. 1995. “Modeling Myopic Decisions: Evidence for Hyperbolic Delay-Discounting with Subjects and Amounts.” *Organizational Behavior and Human Decision Processes*, 64: 22–30.

———. 1996. “Delay-Discounting Probabilistic Rewards: Rates Decrease as Amounts Increase.” *Psychonomic Bulletin and Review*, 3(1): 100–104.

Kirby, Kris N., Nancy M. Petry, and Warren Bickel. 1999. “Heroin Addicts Have Higher Discount Rates for Delayed Rewards than Non-Drug-Using Controls.” *Journal of Experimental Psychology: General*, 128(1): 78–87.

Koomey, Jonathan G., and Alan H. Sanstad. 1994. “Technical Evidence for Assessing the Performance of Markets Affecting Energy Efficiency.” *Energy Policy* 22(10): 826–32.

Koopmans, Tjalling C. 1960. “Stationary Ordinal Utility and Impatience.” *Econometrica*, 28: 287–309.

———. 1967. “Objectives, Constraints, and Outcomes in Optimal Growth Models.” *Econometrica*, 35(1): 1–15.

Koopmans, Tjalling C., Peter A. Diamond, and Richard E. Williamson. 1964. “Stationary Utility and Time Perspective.” *Econometrica*, 32: 82–100.

Koszegi, Botond. 2001. “Who Has Anticipatory Feelings?” Working paper, University of California, Berkeley.

Kroll, Yoram, Haim Levy, and Amnon Rapoport. 1988. “Experimental Tests of the Separation Theorem and the Capital Asset Pricing Model.” *American Economic Review*, 78: 500–19.

Laibson, David. 1994. “Essays in Hyperbolic Discounting.” Ph.D. diss., MIT.

———. 1997. “Golden Eggs and Hyperbolic Discounting.” *Quarterly Journal of Economics*, 112: 443–77.

———. 1998. “Life-Cycle Consumption and Hyperbolic Discount Functions.” *European Economic Review*, 42: 861–71.

———. 2001. “A Cue-Theory of Consumption.” *Quarterly Journal of Economics*, 116: 81–119.

Laibson, David, Andrea Repetto, and Jeremy Tobacman. 1998. “Self-Control and Saving for Retirement.” *Brookings Papers on Economic Activity*, 1: 91–196.

Lancaster, K. J. 1963. “An Axiomatic Theory of Consumer Time Preference.” *International Economic Review*, 4: 221–31.

Lawrence, Emily. 1991. “Poverty and the Rate of Time Preference: Evidence from Panel Data.” *Journal of Political Economy*, 119: 54–77.

LeDoux, Joseph E. 1996. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster.

Loewenstein, George. 1987. “Anticipation and the Valuation of Delayed Consumption.” *Economy Journal*, 97: 666–84.

———. 1988. “Frames of Mind in Intertemporal Choice.” *Management Science*, 34: 200–14.

———. 1996. “Out of Control: Visceral Influences on Behavior.” *Organizational Behavior and Human Decision Processes*, 65: 272–92.

———. 1999. “A Visceral Account of Addiction.” In *Getting Hooked: Rationality and Addiction*, edited by Jon Elster and Ole-Jorgen Skog. Cambridge: Cambridge University Press.

———. 2000a. “Willpower: A Decision-Theorist’s Perspective.” *Law Philosophy*, 19: 51–76.

———. 2000b. “Emotions in Economic Theory and Economic Behavior.” *American Economic Review Papers Proceedings*, 90: 426–32.

Loewenstein, George, Ted O’Donoghue, and Matthew Rabin. 2000. “Projection Bias in the Prediction of Future Utility.” Center for Analytic Economics Working Paper #02-11, Cornell University.

Loewenstein, George, and Drazen Prelec. 1991. “Negative Time Preference.” *American Economic Review*, 81: 347–52.

———. 1992. “Anomalies in Intertemporal Choice: Evidence and an Interpretation.” *Quarterly Journal of Economics*, 107(2): 573–97.

———. 1993. “Preferences for Sequences of Outcomes.” *Psychological Review*, 100(1): 91–108.

Loewenstein, George, and Nachum Sicherman. 1991. “Do Workers Prefer Increasing Wage Profiles?” *Journal of Labor Economics*, 9(1): 67–84.

Loewenstein, George, Roberto Weber, Janine Flory, Stephen Manuck, and Matthew Muldoon. 2001. “Dimensions of Time Discounting.” Paper presented at Conference on Survey Research on Household Expectations and Preferences. Ann Arbor, MI., November 2–3, 2001.

MacCrimmon, Kenneth R., and Donald A. Wehrung. 1990. “Characteristics of Risk-Taking Executives.” *Management Science*, 36(4): 422–35.

MacKeigan, L. D., L. N. Larson, J. R. Draugalis, J. L. Bootman, and L. R. Burns. 1993. “Time Preference for Health Gains vs. Health Losses.” *Pharmacoeconomics*, 3(5): 374–86.

Madden, Gregory J., Nancy M. Petry, Gary J. Badger, and Warren Bickel. 1997. “Impulsive and Self-Control Choices in Opioid-Dependent Patients and Non-Drug-Using Control Participants: Drug and Monetary Rewards.” *Experimental and Clinical Psychopharmacology*, 5(3): 256–62.

Maital, Shloma, and Sharon Maital. 1978. “Time Preference, Delay of Gratification, and Intergenerational Transmission of Economic Inequality: A Behavioral Theory of Income Distribution.” In *Essays in Labor Market Analysis*, edited by Orley Ashenfelter and Wallace Oates. New York: Wiley.

Martin, John L. 2001. “The Authoritarian Personality, 50 Years Later: What Lessons Are There for Political Psychology?” *Political Psychology*, 22(1): 1–26.

Mazur, James E. 1987. “An Adjustment Procedure for Studying Delayed Reinforcement.” In *The Effect of Delay and Intervening Events on Reinforcement Value*, edited by Michael L. Commons, James E. Mazur, John A. Nevin, and Howard Rachlin. Hillsdale, N.J.: Erlbaum.

Meyer, Richard F. 1976. “Preferences Over Time.” In *Decisions with Multiple Objectives*, edited by Ralph Keeney and Howard Raiffa. New York: Wiley.

Millar, Andrew, and Douglas Navarick. 1984. “Self-Control and Choice in Humans: Effects of Video Game Playing as a Positive Reinforcer.” *Learning and Motivation*, 15: 203–18.

Mischel, Walter, Joan Grusec, and John C. Masters. 1969. “Effects of Expected Delay Time on Subjective Value of Rewards and Punishments.” *Journal of Personality and Social Psychology*, 11(4): 363–73.

Mischel, Walter, Yuichi Shoda, and Philip K. Peake. 1988. "The Nature of Adolescent Competencies Predicted by Preschool Delay of Gratification." *Journal of Personality and Social Psychology*, 54(4): 687–96.

Moore, Michael J., and W. Kip Viscusi. 1988. "The Quantity-Adjusted Value of Life." *Economic Inquiry*, 26(3): 369–88.

———. 1990a. "Discounting Environmental Health Risks: New Evidence and Policy Implications." *Journal of Environmental Economics and Management*, 18: S51–S62.

———. 1990b. "Models for Estimating Discount Rates for Long-Term Health Risks Using Labor Market Data." *Journal of Risk Uncertainty*, 3: 381–401.

Munasinghe, Lalith, and Nachum Sicherman. 2000. "Why Do Dancers Smoke? Time Preference, Occupational Choice, and Wage Growth." Working paper, Columbia University and Barnard College, New York.

Murphy, Thomas J., and Alan S. De Wolfe. 1986. "Future Time Perspective in Alcoholics, Process and Reactive Schizophrenics, and Normals." *International Journal of Addictions*, 20: 1815–22.

Myerson, Joel, and Leonard Green. 1995. "Discounting of Delayed Rewards: Models of Individual Choice." *Journal of Experimental Analysis and Behavior*, 64: 263–76.

Nisan, Mordecai, and Abram Minkowich. 1973. "The Effect of Expected Temporal Distance on Risk Taking." *Journal of Personality and Social Psychology*, 25(3): 375–80.

Nyhus, E. K. 1995. "Item and Non Item-Specific Sources of Variance in Subjective Discount Rates. A Cross Sectional Study." Paper presented at the fifteenth Conference on Subjective Probability, Utility and Decision Making, Jerusalem.

O'Donoghue, Ted, and Matthew Rabin. 1999a. "Addiction and Self-Control." In *Addiction: Entries and Exits*, edited by Jon Elster. New York: Russell Sage.

———. 1999b. "Doing It Now or Later." *American Economic Review* 89(1): 103–24.

———. 1999c. "Incentives for Procrastinators." *Quarterly Journal of Economics*, 114(3): 769–816.

———. 1999d. "Procrastination in Preparing for Retirement." In *Behavioral Dimensions of Retirement Economics*, edited by Henry Aaron. New York: Brookings Institution and Russell Sage.

———. 2000. "Addiction and Present-Biased Preferences." Cornell University, Ithaca, and University of California, Berkeley. Center for Analytic Economics Working paper #02-10, Cornell University.

———. 2001. "Choice and Procrastination." *Quarterly Journal of Economics*, 116(1): 121–60.

Olson, Mancur, and Martin J. Bailey. 1981. "Positive Time Preference." *Journal of Political Economy*, 89(1): 1–25.

Orphanides, Athanasios, and David Zervos. 1995. "Rational Addiction with Learning and Regret." *Journal of Political Economy*, 103(4): 739–58.

Parfit, Derek. 1971. "Personal Identity." *Philosophical Review*, 80(1): 3–27.

———. 1976. "Lewis, Perry, and What Matters." In *The Identities of Persons*, edited by Amelie O. Rorty. Berkeley: University of California Press.

———. 1982. "Personal Identity and Rationality." *Synthese*, 53: 227–41.

Peleg, Bezalel, and Menahem E. Yaari. 1973. "On the Existence of a Consistent Course of Action When Tastes Are Changing." *Review of Economic Studies*, 40(3): 391–401.

Pender, John L. 1996. "Discount Rates and Credit Markets: Theory and Evidence from Rural India." *Journal of Developmental Economy*, 50(2): 257–96.

Petry, Nancy M., Warren Bickel, and Martha M. Arnett. 1998. "Shortened Time Horizons and Insensitivity to Future Consequences in Heroin Addicts." *Addiction*, 93: 729–38.

Phelps, E. S., and Robert Pollak. 1968. "On Second-Best National Saving and Game-Equilibrium Growth." *Review of Economic Studies*, 35: 185–99.

Pigou, Arthur C. 1920. *The Economics of Welfare*. London: Macmillan.

Pollak, Robert A. 1968. "Consistent Planning." *Review of Economic Studies*, 35: 201–8.

———. 1970. "Habit Formation and Dynamic Demand Functions." *Journal of Political Economy*, 78(4): 745–63.

Prelec, Drazen, and George Loewenstein. 1998. "The Red and the Black: Mental Accounting of Savings and Debt." *Marketing Science*, 17(1): 4–28.

Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica*, 68(5): 1281–92.

Rabin, Matthew, and Richard H. Thaler. 2001. "Anomalies: Risk Aversion." *Journal of Economic Perspectives*, 15(1): 219–32.

Rachlin, Howard, Andres Raineri, and David Cross. 1991. "Subjective Probability and Delay," *Journal of Experimental Analysis and Behavior*, 55(2): 233–44.

Rae, John. 1905 [1834]. *The Sociological Theory of Capital*. London: Macmillan.

Raineri, Andres, and Howard Rachlin. 1993. "The Effect of Temporal Constraints on the Value of Money and Other Commodities." *Journal of Behavioral Decision Making*, 6: 77–94.

Read, Daniel. 2001. "Is Time-Discounting Hyperbolic or Subadditive?" *Journal of Risk Uncertainty*, 23: 5–32.

Read, Daniel, George F. Loewenstein, and Matthew Rabin. 1999. "Choice Bracketing." *Journal of Risk Uncertainty*, 19: 171–97.

Redelmeier, Daniel A., and Daniel N. Heller. 1993. "Time Preference in Medical Decision Making and Cost-Effectiveness Analysis." *Medical Decision Making*, 13(3): 212–17.

Roelofsma, Peter. 1994. "Intertemporal Choice." Free University, Amsterdam.

Ross, Jr., William T., and Itamar Simonson. 1991. "Evaluations of Pairs of Experiences: A Preference for Happy Endings." *Journal of Behavioral Decision Making*, 4: 155–61.

Roth, Alvin E., and J. Keith Murnighan. 1982. "The Role of Information in Bargaining: An Experimental Study." *Econometrica*, 50(5): 1123–42.

Rubinstein, Ariel. 2000. "Is It 'Economics and Psychology'? The Case of Hyperbolic Discounting." Tel Aviv University and Princeton University.

Ruderman, H., M. D. Levine, and J. E. McMahon. 1987. "The Behavior of the Market for Energy Efficiency in Residential Appliances Including Heating and Cooling Equipment." *Energy Journal*, 8(1): 101–24.

Ryder, Harl E., and Geoffrey M. Heal. 1973. "Optimal Growth with Intertemporally Dependent Preferences." *Review of Economic Studies*, 40: 1–33.

Samuelson, Paul. 1937. "A Note on Measurement of Utility." *Review of Economic Studies*, 4: 155–61.

———. 1952. "Probability, Utility, and the Independence Axiom." *Econometrica*, 20(4): 670–78.

Schelling, Thomas C. 1984. "Self-Command in Practice, in Policy, and in a Theory of Rational Choice." *American Economic Review*, 74(2): 1–11.

Senior, N. W. 1836. *An Outline of the Science of Political Economy*. London: Clowes and Sons.

Shea, John. 1995a. "Myopia, Liquidity Constraints, and Aggregate Consumption." *Journal of Money, Credit, Banking*, 27(3): 798–805.

———. 1995b. "Union Contracts and the Life-Cycle/Permanent-Income Hypothesis." *American Economic Review*, 85(1): 186–200.

Shelley, Marjorie K. 1993. “Outcome Signs, Question Frames and Discount Rates.” *Management Science* 39: 806–15.

———. 1994. “Gain/Loss Asymmetry in Risky Intertemporal Choice.” *Organizational Behavior and Human Decision Processes*, 59: 124–59.

Shelley, Marjorie K., and Thomas C. Omer. 1996. “Intertemporal Framing Issues in Management Compensation.” *Organizational Behavior and Human Decision Processes*, 66(1): 42–58.

Shoda, Yuichi, Walter Mischel, and Philip K. Peake. 1990. “Predicting Adolescent Cognitive and Self-Regulatory Competencies from Preschool Delay of Gratification.” *Developmental Psychology*, 26(6): 978–86.

Solnick, Jay, Catherine Kannenberg, David Eckerman, and Marcus Waller. 1980. “An Experimental Analysis of Impulsivity and Impulse Control in Humans.” *Learning and Motivation*, 11: 61–77.

Solow, Robert M. 1974. “Intergenerational Equity and Exhaustible Resources.” *Review of Economic Studies*, 41: Symposium on the Economics of Exhaustible Resources, 29–45.

Spence, Michael, and Richard Zeckhauser. 1972. “The Effect of Timing of Consumption Decisions and Resolution of Lotteries on Choice of Lotteries.” *Econometrica*, 40(2): 401–403.

Starmer, Chris. 2000. “Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk.” *Journal of Economic Literature*, 38(2): 332–82.

Sternberg, Robert J. 1985. *Beyond IQ: A Triarchic Theory of Human Intelligence*. New York: Cambridge University Press.

Stevenson, Mary Kay. 1992. “The Impact of Temporal Context and Risk on the Judged Value of Future Outcomes.” *Organizational Behavior and Human Decision Processes*, 52(3): 455–91.

Strotz, Robert H. 1955–56. “Myopia and Inconsistency in Dynamic Utility Maximization.” *Review of Economic Studies*, 23(3): 165–80.

Suranovic, Steven, Robert Goldfarb, and Thomas C. Leonard. 1999. “An Economic Theory of Cigarette Addiction.” *Journal of Health Economics*, 18(1): 1–29.

Thaler, Richard H. 1981. “Some Empirical Evidence on Dynamic Inconsistency.” *Economic Letters*, 8: 201–207.

———. 1985. “Mental Accounting and Consumer Choice.” *Management Science*, 4: 199–214.

———. 1999. “Mental Accounting Matters.” *Journal of Behavioral Decision Making*, 12: 183–206.

Thaler, Richard H., and Hersh M. Shefrin. 1981. “An Economic Theory of Self-Control.” *Journal of Political Economy*, 89(2): 392–410.

Tversky, Amos, and Daniel Kahneman. 1983. “Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment.” *Psychological Review*, 90: 293–315.

———. 1991. “Loss Aversion in Riskless Choice: A Reference Dependent Model.” *Quarterly Journal of Economics*, 106: 1039–61.

Tversky, Amos, and Derek J. Koehler. 1994. “Support Theory: Nonextensional Representation of Subjective Probability.” *Psychological Review*, 101(4): 547–67.

van der Pol, Marjon M., and John A. Cairns. 1999. “Individual Time Preferences for Own Health: Application of a Dichotomous Choice Question with Follow Up.” *Applied Economic Letters*, 6(10): 649–54.

———. 2001. “Estimating Time Preferences for Health Using Discrete Choice Experiments.” *Social Science and Medicine*, 52: 1459–70.

Varey, Carol A., and Daniel Kahneman. 1992. "Experiences Extended Across Time: Evaluation of Moments and Episodes." *Journal of Behavioral Decision Making*, 5(3): 169–85.

Viscusi, W. Kip, and Michael J. Moore. 1989. "Rates of Time Preference and Valuation of the Duration of Life." *Journal of Public Economics*, 38(3): 297–317.

Wahlund, Richard, and Jonas Gunnarsson. 1996. "Mental Discounting and Financial Strategies." *Journal of Economic Psychology*, 17(6): 709–30.

Wang, Ruqu. 1997. "The Optimal Consumption and Quitting of Harmful Addictive Goods." Working paper, Queens University, New York.

Warner, John T., and Saul Pleeter. 2001. "The Personal Discount Rate: Evidence from Military Downsizing Programs." *American Economic Review*, 91(1): 33–53.

Whiting, Jennifer. 1986. "Friends and Future Selves." *Philosophical Review*, 95(4): 547–80.

Winston, Gordon C. 1980. "Addiction and Backsliding: A Theory of Compulsive Consumption." *Journal of Economic Behavioral Organization*, 1: 295–324.

Yates, J. Frank, and Royce A. Watts. 1975. "Preferences for Deferred Losses." *Organizational Behavior and Human Performance*, 13(2): 294–306.

C H A P T E R 7

Doing It Now or Later

TED O'DONOGHUE AND MATTHEW RABIN

PEOPLE are impatient—they like to experience rewards soon and to delay costs until later. Economists almost always capture impatience by assuming that people discount streams of utility over time exponentially. Such preferences are *time-consistent*: A person's relative preference for well-being at an earlier date over a later date is the same no matter when she is asked.

Casual observation, introspection, and psychological research all suggest that the assumption of time consistency is importantly wrong.[1] It ignores the human tendency to grab immediate rewards and to avoid immediate costs in a way that our "long-run selves" do not appreciate. For example, when presented a choice between doing seven hours of an unpleasant activity on April 1 versus eight hours on April 15, if asked on February 1 virtually everyone would prefer the seven hours on April 1. But come April 1, given the same choice, most of us are apt to put off the work until April 15. We call such tendencies *present-biased preferences*: When considering trade-offs between two future moments, present-biased preferences give stronger relative weight to the earlier moment as it gets closer.[2]

In this chapter, we explore the behavioral and welfare implications of present-biased preferences in a simple model where a person must engage in an activity

We thank Steven Blatt, Erik Eyster, and Clara Wang for useful research assistance, and Steven Blatt, Erik Eyster, David Laibson, two anonymous referees, and seminar participants at the University of California Berkeley, Northwestern University, the Russell Sage Foundation, Columbia University, Yale University, Harvard University, MIT, the University of Wisconsin, Cornell University, Arizona State University, Santa Clara University, Texas A&M University, and the University of Chicago for helpful comments. For financial support, we thank the National Science Foundation (Grant No. 9709485). O'Donoghue thanks the Alfred P. Sloan Foundation and the Math Center at Northwestern University, and Rabin thanks the Alfred P. Sloan and Russell Sage Foundations.

[1] Loewenstein (1992) reviews how the economics profession evolved from perceiving exponential discounting as a useful, ad hoc approximation of intertemporal-choice behavior, to perceiving it as a fundamental axiom of (rational) human behavior. For some recent discussions of empirical evidence of time inconsistency, see Thaler (1991) and Thaler and Loewenstein (1992).

[2] Many researchers have studied time-inconsistent preferences. A small set of economists have over the years proposed formal, general models of time-inconsistent preferences. See, for instance, Strotz (1956), Phelps and Pollak (1968), Pollak (1968), Peleg and Yaari (1973), Yaari (1977), and Goldman (1979, 1980). Other researchers have posited a specific functional form, hyperbolic discounting, to account for observed tendencies for immediate gratification [see Chung and Herrnstein (1967), Ainslie and Herrnstein (1981), Ainslie (1991, 1992), Ainslie and Haslam (1992b), and Loewenstein and Prelec (1992)]. We have contrived the term "present-biased preferences" as a more descriptive term for the underlying human characteristic that hyperbolic discounting represents.

exactly once during some length of time. This simple model encompasses an important class of situations, and also allows us to lay bare some basic principles that might apply more generally to formal models of time-inconsistent preferences.

Our analysis emphasizes two sets of distinctions. The first distinction is whether choices involve *immediate costs*—where the costs of an action are immediate but any rewards are delayed—or *immediate rewards*—where the benefits of an action are immediate but any costs are delayed. By exploring these two different settings under the rubric of present-biased preferences, we unify the investigation of phenomena (e.g., procrastination and overeating) that have often been explored separately, but which clearly come from the same underlying propensity for immediate gratification.[3]

The second distinction is whether people are *sophisticated*, and foresee that they will have self-control problems in the future, or are naïve and do not foresee these self-control problems. By explicitly comparing these competing assumptions—each of which has received attention in the economics literature—we hope to delineate which predictions come from present-biased preferences per se, and which come from these assumptions about foresight.[4]

In section 1, we further motivate and formally define a simplified form of present biased preferences [originally proposed by Phelps and Pollak (1968) and later employed by Laibson (1994)] that we study in this paper: Relative to time-consistent preferences, a person always gives extra weight to well-being now over any future moment but weighs all future moments equally. In section 2, we set up our model of a one-time activity. We suppose that a person must engage in an activity exactly once during some length of time. Importantly, at each moment the person can choose only whether or not to do it now, and cannot choose when later she will do it. Within this scenario, we consider a general class of reward and cost schedules for completing the activity.

Section 3 explores the behavioral implications of present-biased preferences in our model. We present two simple results characterizing how behavior depends on whether rewards or costs are immediate, and on whether people are sophisticated or naïve. The *present-bias effect* characterizes the direct implications of present-biased preferences: You *procrastinate*—wait when you should do it—if actions involve immediate costs (writing a paper), and *preproperate*—do it when you should wait—if actions involve immediate rewards (seeing a movie). Naïve people are

[3] Throughout this chapter, our emphasis is on impulsive choice driven by a tendency to overweight rewards and costs that are in close temporal proximity. But there are clearly other aspects of impulsive choice as well: People also tend to overweight rewards and costs that are in close spatial proximity, and more generally are attentive to rewards and costs that are salient (see Loewenstein, 1996).

[4] Strotz (1956) and Pollak (1968), two of the seminal papers on time-inconsistent preferences, carefully lay out these two assumptions, but do not much consider the implications of one versus the other. More recent papers have assumed either one or the other, without attempting to justify the choice on behavioral grounds. For instance, George A. Akerlof (1991) assumes naive beliefs, while David Laibson (1994, 1995, 1997) and Carolyn Fischer (1997) assume sophisticated beliefs. Each paper states its assumption about beliefs used [and Akerlof (1991) posits that his main welfare finding depends on his assumption of naive beliefs], but conspicuously does not argue why its assumption is correct.

influenced solely by the present-bias effect. The *sophistication effect* characterizes the direct implications of sophistication versus naïvete: A sophisticated person does the activity sooner than does a naive person with the same preferences, irrespective of whether rewards or costs are immediate. Intuitively, a sophisticated person is correctly pessimistic about her future behavior—a naïve person believes she will behave herself in the future while a sophisticated person knows she may not. As a result, waiting always seems less attractive for a sophisticated person. Although the direction is the same, the sophistication effect has very different connotations for immediate costs versus immediate rewards. When costs are immediate, sophistication mitigates the tendency to procrastinate. (And in fact, the sophistication effect can outweigh the present-bias effect so that a sophisticated person may perform an onerous activity before she would if she had no self-control problem.) When rewards are immediate, on the other hand, sophistication exacerbates the tendency to preproperate.

In section 4 we turn to the welfare results.[5] Again, the two distinctions—immediate costs vs. immediate rewards and sophistication vs. naivete—are crucial. When costs are immediate, a person is always better off with sophisticated beliefs than with naïve beliefs. Naïvete can lead you to repeatedly procrastinate an unpleasant activity under the incorrect belief that you will do it tomorrow, while sophistication means you know exactly how costly delay would be. In fact, even with an arbitrarily small bias for the present, for immediate costs naive people can experience severe welfare losses, while the welfare loss from a small present bias is small if you are sophisticated. When rewards are immediate, however, a person can be better off with naïve beliefs. In this case, people with present-biased preferences tend to do the activity when they should wait. Naïvete helps motivate you to wait because you overestimate the benefits of waiting. Sophistication makes you (properly) skeptical of future behavior, so you are more tempted to grab today's immediate reward. This can lead to "unwinding" similar to that in the finitely repeated prisoner's dilemma: In the end, you will give in to temptation and grab a reward too soon; because you realize this, near the end you will cave in a little sooner than if you thought you would resist temptation in the end; realizing this, you will cave in a little sooner, etc. As a result, for immediate rewards it is sophisticated people who can experience severe welfare losses with an arbitrarily small present bias, while the welfare loss from a small present bias is small if you are naive.

Researchers looking for empirical proof of time-inconsistent preferences often explore the use of self-limiting "commitment devices" (e.g., Christmas clubs, fat farms), because such devices represent "smoking guns" that cannot be explained by any time-consistent preferences. We show in section 5 that even within our simple

[5] Welfare comparisons for people with time-inconsistent preferences are in principle problematic; the very premise of the model is that a person's preferences disagree at different times, so that a change in behavior may make some selves better off while making other selves worse off. We feel the natural perspective in most situations is the "long-run perspective"—what you would wish now (if you were fully informed) about your profile of future behavior. However, few of our comparisons rely on this perspective, and most of our welfare comparisons can be roughly conceived of as "Pareto comparisons," where one outcome is better than another from all of a person's vantage points.

setting, certain behaviors induced by present-biased preferences are inconsistent with any time-consistent preferences. Hence, we illustrate that smoking guns need not involve external commitment devices. Furthermore, while previous literature has focused on smoking guns for sophisticated people, we show that smoking guns exist for naive people as well.

Although many of the specific results described above are special to our one-activity model, these results illustrate some more general intuitions. To begin the process of generalizing our model, in section 6 we present an extension where, rather than being performed exactly once, the activity must be performed more than once during some length of time. In section 7, we discuss more broadly (and less formally) what our model suggests about general implications of self-control problems, and describe how some of these implications might play out in specific economic contexts, such as saving and addiction. We then conclude with a discussion of some lessons to take away from our analysis, both for why it is important that economists start to study self-control problems, and for how we should go about doing so.

1. Present-Biased Preferences

Let u_t be a person's *instantaneous utility* in period t. A person in period t cares not only about her present instantaneous utility, but also about her future instantaneous utilities. We let $U^t(u_t, u_{t+1}, \ldots, u_T)$ represent a person's *intertemporal preferences* from the perspective of period t, where U^t is continuous and increasing in all components.[6] The standard simple model employed by economists is exponential discounting: For all t, $U^t(u_t, u_{t+1}, \ldots, u_T) \equiv \Sigma_{\tau=t}^{T} \delta^T u_\tau$, where $\delta \in (0,1]$ is a "discount factor."

Exponential discounting parsimoniously captures the fact that people are impatient. Yet exponential discounting is more than an innocuous simplification of a more general class of preferences, since it implies that preferences are *time-consistent*: A person's relative preference for well-being at an earlier date over a later date is the same no matter when she is asked. But intertemporal preferences are not time-consistent. People tend to exhibit a specific type of time-inconsistent preferences that we call *present-biased preferences*: When considering trade-offs between two future moments, present-biased preferences give stronger relative weight to the earlier moment as it gets closer.[7]

[6] Note that this formalization is entirely agnostic about what factors appear as arguments in the instantaneous utility function. For instance, while it is common to assume that a person's instantaneous utility u_t depends only on her consumption bundle in period t, our formulation also allows for instantaneous utilities to depend on past consumption (as suggested by Becker and Murphy 1988; Kahneman et al. 1991).

[7] We have contrived the term "presented-biased preferences" to cannote that people's preferences have a bias for the "present" over the "future" (where the "present" is constantly changing). This is merely our term for an array of older models that went under different names. In fact, the (β, δ)-preferences that we will use in this paper are identical to the preferences studied by Laibson (1994), who uses the term "hyperbolic discounting," and are essentially identical to the preferences used in Akerlof (1991),

In this paper, we adopt an elegant simplification for present-biased preferences developed by Phelps and Pollak (1968), and later employed by Laibson (1994, 1995, 1997), Fischer (1997), and O'Donoghue and Rabin (1999). They capture the most basic form of present-biased preferences—a bias for the "present" over the ''future"—with a simple two-parameter model that modifies exponential discounting.

Definition 1. (β, δ)-preferences are preferences that can be represented by

$$\text{For all } t,\ U^t(u_t, u_{t+1}, \ldots, u_T) \equiv \delta^t u_t + \beta \sum_{\gamma=t+1}^{T} \delta^T u_\tau$$

where $0 < \beta, \delta \leq 1$.

In this model, δ represents long-run, time-consistent discounting. The parameter β, on the other hand, represents a "bias for the present"—how you favor now versus later. If $\beta = 1$, then (β, δ)-preferences are simply exponential discounting. But $\beta < 1$ implies present-biased preferences: The person gives more relative weight to period τ in period τ than she did in any period prior to period τ.

Researchers have converged on a simple strategy for modeling time-inconsistent preferences: The person at each point in time is modeled as a separate "agent" who is choosing her current behavior to maximize current preferences, where her future selves will control her future behavior. In such a model, we must ask what a person believes about her future selves' preferences. Strotz (1956) and Pollak (1968) carefully lay out two extreme assumptions. A person could be *sophisticated* and know exactly what her future selves' preferences will be. Or, a person could be naïve and believe her future selves' preferences will be identical to those of her current self, not realizing that as she gets closer to executing decisions her tastes will have changed. We could, of course, also imagine more intermediate assumptions. For instance, a person might be aware that her future selves will have present-biased preferences, but underestimate the degree of the present bias. Except for a brief comment in section 7, we focus in this paper entirely on the two extreme assumptions.

Are people sophisticated or naïve?[8] The use of self-commitment devices, such as alcohol clinics, Christmas clubs, or fat farms, provides evidence of sophistication.[9] Only sophisticated people would want to commit themselves to smaller choice

although Akerlof frames his discussion very differently. For more general definitions of present-biased preferences and related elements of our model, see O'Donoghue and Rabin (1996). For an alternative formulation of the same phenomenon, see Prelec (1990), who uses the term "decreasing impatience."

[8] Most economists modeling time-inconsistent preferences assume sophistication. Indeed, sophistication implies that people have "rational expectations" about future behavior, so it is a natural assumption for economists. Akerlof (1991) uses a variant of the naivete assumption.

[9] The very term "self-control" implies that people are aware that it may be prudent to control their future selves. For analyses of self-control in people, see Ainslie (1974, 1975, 1987, 1992), Schelling (1978, 1984, 1992), Thaler (1980), Thaler and Shefrin (1981), Funder and Block (1989), Hoch and

sets: If you were naïve, you would never worry that your tomorrow self might choose an option you do not like today. Despite the existence of some sophistication, however, it does appear that people underestimate the degree to which their future behavior will not match their current preferences over future behavior. For example, people may repeatedly not have the "will power" to forgo tempting foods or to quit smoking, while predicting that tomorrow they will have this will power. We think there are elements of both sophistication and naïvete in the way people anticipate their own future preferences. In any event, our goal is to clarify the logic of each, and in the process we delineate which predictions come purely from present-biased preferences, and which come from the "sophistication effects" of people being aware of their own time inconsistency.

2. Doing It Once

Suppose there is an activity that a person must perform exactly once, and there are $T < \infty$ periods in which she can do it. Let $\mathbf{v} = (v_1, v_2, \ldots, v_T)$ be the *reward schedule*, and let $\mathbf{c} \equiv (c_1, c_2, \ldots, c_T)$ be the *cost schedule*, where $v_t \geq 0$ and $c_t \geq 0$ for each $t \in \{1, 2, \ldots, T\}$. In each period $t \leq T - 1$, the person must choose either to do it or to wait. If she does the activity in period t, she receives reward v_t but incurs cost c_t, and makes no further choices. If she waits, she then will face the same choice in period $t + 1$. Importantly, if the person waits she cannot commit in period t to when later she will do it. If the person waits until period T, she must do it then.

The reward schedule $\mathbf{v}$ and the cost schedule $\mathbf{c}$ represent rewards and costs as a function of when the person does the activity. However, the person does not necessarily receive the rewards and costs immediately upon completion of the activity. Indeed, we differentiate cases precisely by when rewards and costs are experienced. Some activities, such as writing a paper or mowing the lawn, are unpleasant to perform, but create future benefits. We refer to activities where the cost is incurred immediately while the reward is delayed as activities having *immediate costs*. Other activities, such as seeing a movie or taking a vacation, are pleasurable to perform, but may create future costs. We refer to activities where the reward is received immediately while the cost is delayed as activities having *immediate rewards*.[10]

We analyze these two cases using the (β, δ) preferences outlined in section 1. For simplicity, we assume $\delta = 1$; i.e., we assume that there is no "long-term"

Loewenstein (1991), Ainslie and Haslam (1992a), Glazer and Weiss (1992), Shefrin and Thaler (1992), Wertenbroch (1993), and Laibson (1994, 1995, 1997). Ainslie (1974) explores similar issues with pigeons. As many have emphasized, especially Ainslie (1992) and Watterson (1993, pp. 83–88), a sort of intrapersonal "bargaining" can arise because of the basic disagreements we have with ourselves about when we should do something.

[10] We occasionally make reference to a third case where both rewards and costs are immediate. The fourth case–neither rewards nor costs are immediate—is not of interest because it is equivalent to the case of time consistency, which we study.

discounting.[11] Given $\delta = 1$, without loss of generality we can interpret delayed rewards or costs as being experienced in period $T + 1$. We can then describe a person's intertemporal utility from the perspective of period t of completing the activity in period $\tau \geq t$, which we denote by $U^t(\tau)$.[12]

1. *Immediate Costs*. If a person completes the activity in period τ, then her inter-temporal utility in period $t \leq \tau$ is

$$U^t(\tau) = \begin{cases} \beta v_\tau - c_\tau & \text{if } \tau = t \\ \beta v_\tau - \beta c_\tau & \text{if } \tau > t. \end{cases}$$

2. *Immediate Rewards*. If a person completes the activity in period τ, then her inter-temporal utility in period $t \leq \tau$ is

$$U^t(\tau) = \begin{cases} v_\tau - \beta c_\tau & \text{if } \tau = t \\ \beta v_\tau - \beta c_\tau & \text{if } \tau > t. \end{cases}$$

We will focus in this environment on three types of agents. We refer to people with standard exponential, time-consistent preferences (i.e., $\beta = 1$) as *TC*s. We then focus on two types of people with present-biased preferences (i.e., $\beta < 1$), representing the two extremes discussed in section 1. We call people with sophisticated perceptions *sophisticates*, and people with naïve perceptions *naifs*. Sophisticates and naïfs have identical preferences (throughout we assume they have the same β), and therefore differ only in their perceptions of future preferences.

A person's behavior can be fully described by a *strategy* $\mathbf{s} \equiv (s_1, s_2, \ldots, s_T)$, where $s_t \in [Y, N]$ specifies for period $t \in [1, 2, \ldots, T]$ whether or not to do the activity in period t given she has not yet done it. The strategy $\mathbf{s}$ specifies doing it in period t if $s_t = Y$, and waiting if $s_t = N$. In addition to specifying when the person will actually complete the activity, a strategy also specifies what the person "would" do in periods after she has already done it; e.g., if $s_t = Y$, we still specify $s_{t'}$ for all $t' > t$. This feature will prove useful in our analysis. Since the person must do it in period T if she has not yet done it, without loss of generality we require $s_T = Y$.

To describe behavior given our assumptions, we define a "solution concept": A *perception-perfect strategy* is a strategy that in all periods (even those after the activity is performed) a person chooses the optimal action given her current preferences

[11] The results are easily generalized to $\delta < 1$. Suppose the "true" reward schedule is $\pi = (\pi_1, \pi_2, \ldots, \pi_\tau)$, the "true" cost schedule is $\phi = (\phi_1, \phi_2, \ldots, \phi_\tau)$, and $\delta < 1$. If, for instance, costs are immediate and rewards are received in period $T + 1$, then if we let $v_t = \delta^{\tau+1}\pi_t$ and $c_t = \delta^t\phi_t$ for each t, doing the analysis with $\mathbf{v}$, $\mathbf{c}$, and no discounting is identical to doing the analysis with π, ϕ, and δ.

[12] This formulation normalizes the instantaneous utility from not completing the activity to be zero. For instance, when costs are immediate and rewards are received in period $T + 1$, we are assuming that if the person does the activity in period τ, the instantaneous utilities are $u_r = -c_r$, $u_{l+i} = v_i$, and $u_t = 0$ for all $t \notin \{\tau, v + 1\}$. This assumption is purely for convenience. In particular, for any $\bar{u}$ we would get identical results if we normalize the utility from not doing the activity to be $\bar{u}$, or if we normalize the utility from completing the activity to be $\bar{u}$.

and her perceptions of future behavior. Rather than give a general formal definition, we simply define a perception-perfect strategy for each of the three types of agents that we consider. Definition 2 describes a perception-perfect strategy for TCs. Reflecting the fact that TCs do not have a self-control problem, definition 2 says that in any period. TCs will complete the activity if and only if it is the optimal period of those remaining given her current preferences.

Definiton 2. A perception-perfect strategy for *TC*s is a strategy $\mathbf{s}^{tc} \equiv (s_1^{tc}, s_2^{tc}, \ldots, s_T^{tc})$ satisfies for all $t < T$ $s_t^{tc} = Y$ if and only if $U^t(t) \geq U^t(\tau)$ for all $\tau > t$.

Naïfs have present-biased preferences (since $\beta < 1$), but naïfs believe that they are time-consistent. As a result, the decision process for naïfs is identical to that for TCs (although naïfs have different preferences). Definition 3 says that in any period, naïfs will complete the activity if and only if it is the optimal period of those remaining given her current preferences.

Definition 3. A perception-perfect strategy for naïfs is a strategy $\mathbf{s}^n \equiv (s_1^n, s_2^n, \ldots, s_T^n)$ that satisfies for all $t < T$ $s_t^n = Y$ if and only if $U^t(t) \geq U^t(\tau)$ for all $\tau > t$.

Although naïfs and TCs have essentially the same decision process, it is important to realize that naïfs have incorrect perceptions about future behavior, and therefore may plan to behave one way but in fact behave differently. With (β, δ)-preferences, these incorrect perceptions take a convenient form: At all times, naifs believe that if they wait they will behave like TCs in the future.

Sophisticates also have present-biased preferences and a self-control problem. But unlike naïfs, sophisticates know they will have self-control problems in the future, and therefore correctly predict future behavior. Definition 4 says that in period *t*, sophisticates calculate when their future selves will complete the activity if they wait now, and then do the activity now if and only if given their current preferences doing it now is preferred to waiting for their future selves to do it.

Definition 4. A perception-perfect strategy for sophisticates is a strategy $\mathbf{s}^s \equiv (s_1^s, s_2^s, \ldots, s_T^s)$ that satisfies for all $t < T$ $s_t^s = Y$ if and only if $U^t(t) \geq U^t(\tau')$ where $\tau' \equiv \min_{\tau > t}\{\tau \mid s_\tau^s = Y\}$.

Note that in definitions 2, 3, and 4 we have assumed that people do it when indifferent, which implies that there is a unique perception-perfect strategy for each type. In addition, this assumption implies that a perception-perfect strategy must be a pure strategy. For generic values of **v**, **c**, and β, nobody will ever be indifferent, so these assumptions are irrelevant. In nongeneric games, more general definitions could lead to additional equilibria. For sophisticates, a perception-perfect strategy is the identical solution concept to that used by Strotz (1956), Pollak (1968), Laibson (1994, 1995, 1997), and others. For naïfs, it is essentially the same solution concept as those used by Pollak (1968) and Akerlof (1991).

It will be useful in the analysis of this model to have notation for when a person will actually complete the activity (i.e., the outcome): Given the perception-perfect strategies $\mathbf{s}^{tc}$, $\mathbf{s}^{s}$, and $\mathbf{s}^{n}$, we let τ_{tc}, τ_s, and τ_n be the periods in which each of the three types of agents do the activity. That is, given $a \in \{tc, s, n\}$, $\tau_a \equiv \min_t \{t \mid s_t^a = Y\}$.

3. Behavior

In this section, we compare the behavior of TCs, naïfs, and sophisticates who have identical long-run preferences. Comparing naifs or sophisticates to TCs reflects how people with present-biased preferences behave from a long-run perspective; and comparing sophisticates to naifs reflects the implications of sophistication about self-control problems.

We begin by analyzing in some detail a pair of related examples to illustrate the intuitions behind many of the results. Consider the following scenario: Suppose you usually go to the movies on Saturdays, and the schedule at the local cinema consists of a mediocre movie this week, a good movie next week, a great movie in two weeks, and (best of all) a Johnny Depp movie in three weeks. Now suppose you must complete a report for work within four weeks, and to do so you must skip the movie on one of the next four Saturdays. When do you complete the report?

The activity you must do exactly once is writing the report. The reward from doing the report is received at work in the future. We will assume the reward is independent of when you complete the report, and denote it by $\overline{v}$. The cost of doing the report on a given Saturday—not seeing the movie shown that day—is experienced immediately. Letting valuations of the mediocre, good, great, and Depp movies be 3, 5, 8, and 13, we formalize this situation in the following example, where we present both the parameters of the example and the perception-perfect strategy for each type of agent.

Example 1. Suppose costs are immediate, $T = 4$, and $\beta = 1/2$ for naifs and sophisticates. Let $\mathbf{v} = (\overline{v}, \overline{v}, \overline{v}, \overline{v})$ and $\mathbf{c} = (3, 5, 8, 13)$.

$\mathbf{s}^{tc} = (Y, Y, Y, Y)$, so TCs do the report in period $\tau_{tc} = 1$.
$\mathbf{s}^{n} = (N, N, N, Y)$, so naifs do the report in period $\tau_n = 4$.
$\mathbf{s}^{s} = (N, Y, N, Y)$, so sophisticates do the report in period $\tau_s = 2$.

TCs do the report on the first Saturday, skipping the mediocre movie. TCs always do the activity in the period t that maximizes $\theta_t - c_t$. Since example 1 has a stationary reward schedule, TCs do the report in the period with the minimum cost.

Naïfs procrastinate until the last Saturday, forcing themselves to skip the Depp movie. On the first Saturday, naifs give in to their self-control problem and see the mediocre movie because they believe they will skip the good movie in week 2 and still be able to see the great movie and the Depp movie. The period-1 naïf prefers incurring a cost of 5 next week as opposed to a cost of 3 now. However, when the second Saturday arrives, naïfs again give in to their self-control problem and see the good movie, now believing they will skip the great movie in week 3 and still get to see the Depp movie. Finally, when the third Saturday arrives, naïfs have self-control problems for a third time and see the great movie, forcing themselves to miss the Depp movie. This example demonstrates a typical problem for naïfs when costs are immediate: They incorrectly predict that they will not procrastinate in the future, and consequently underestimate the cost of procrastinating now.

Sophisticates procrastinate one week, but they do the report on the second Saturday, skipping the good movie and enabling themselves to see the great movie and the Depp movie. The period-1 sophisticate correctly predicts that he would have self-control problems on the third Saturday and see the great movie. However, the period-1 sophisticate also correctly predicts that knowing about period-3 self-control problems will induce him to do the report on the second Saturday. Hence, the period-1 sophisticate can safely procrastinate and see the mediocre movie: Example 1 illustrates typical behavior for sophisticates when costs are immediate. Although sophisticates have a tendency to procrastinate (they do not write the report right away, which their long-run selves prefer), perfect foresight can mitigate this problem because sophisticates will do it now when they (correctly) foresee costly procrastination in the future.

Example 1 illustrates an intuition expressed by Strotz (1956) and Akerlof (1991) that sophistication is "good" because it helps overcome self-control problems. As in Akerlof's (1991) procrastination example, naïfs repeatedly put off an activity because they believe they will do it tomorrow. Akerlof intuits that sophistication could overcome this problem, and example 1 demonstrates this intuition.

However, this intuition may not hold when rewards are immediate. Consider a similar scenario: Suppose you have a coupon to see one movie over the next four Saturdays, and your allowance is such that you cannot afford to pay for a movie. The schedule at the local cinema is the same as for the above example—a mediocre movie this week, a good movie next week, a great movie in two weeks, and (best of all) a Johnny Depp movie in three weeks. Which movie do you see?

Now, the activity you must do exactly once is going to a movie, and the reward, seeing the movie, is experienced immediately.[13] Using the same payoffs for seeing a movie as in example 1, we have the following formalization.

Example 2. Suppose rewards are immediate, $T = 4$, and $\beta = 1/2$ for naïfs and sophisticates. Let $\mathbf{v} = (3, 5, 8, 13)$ and $\mathbf{c} = (0, 0, 0, 0)$.

$\mathbf{s}^{tc} = (N, N, N, Y)$, so TCs see the movie in period $\tau_{tc} = 4$.
$\mathbf{s}^{n} = (N, N, Y, Y)$, so naifs see the movie in period $\tau_{n} = 3$.
$\mathbf{s}^{s} = (Y, Y, Y, Y)$, so sophisticates see the movie in period $\tau_{s} = 1$.

TCs wait and see the Depp movie since it yields the highest reward. Naïfs see merely the great movie. On the first two Saturdays, naïfs skip the mediocre and

[13] That seeing a movie is a "cost" in example 1 and a "reward" in example 2 reflects that the rewards and costs are defined with respect to the activity being done once.

good movies incorrectly believing they will wait to see the Depp movie. However, on the third Saturday, they give in to self-control problems and see the great movie. For activities with immediate rewards, the self-control problem leads naïfs to do the activity too soon.

Sophisticates have even worse self-control problems in this situation. They see merely the mediocre movie because of an unwinding similar to that in the finitely repeated prisoner's dilemma. The period-2 sophisticate would choose to see the good movie because he correctly predicts that he would give in to self-control problems on the third Saturday, and see merely the great movie rather than the Depp movie. The period-1 sophisticate correctly predicts this reasoning and behavior by his period-2 self. Hence, the period-1 sophisticate realizes that he will see merely the good movie if he waits, so he concludes he might as well see the mediocre movie now. This example demonstrates a typical problem for sophisticates when rewards are immediate: Knowing about future self-control problems can lead you to give in to them today, because you realize you will give in to them tomorrow.[14]

We now present some propositions that characterize present-biased behavior more generally. We refer to the most basic intuition concerning how present-biased preferences affect behavior as the *present-bias effect*:[15] When costs are immediate people with present-biased preferences tend to *procrastinate*—wait when they should do it—while when rewards are immediate they tend to *preproperate*—do it when they should wait.[16] For immediate costs, they wait in periods where they should do it because they want to avoid the immediate cost. For immediate rewards, they do it in periods where they should wait because they want the immediate reward now. Proposition 1 captures that naïfs are influenced solely by the present-bias effect—for immediate costs naïfs always procrastinate, and for immediate rewards naifs always preproperate.[17]

Proposition 1. (1) If costs are immediate, then $\tau \geq \tau_{tc}$. (2) If rewards are immediate, then $\tau_n \leq \tau_{tc}$.

[14] The example also shows why sophisticates would like ways to "commit" the behavior of their future selves, as discussed by many researchers: If the period-1 sophisticate could commit himself to seeing the Depp or great movie, he would do so—even given his taste for immediate rewards. Note that with a reasonable assumption that a person does not bind himself when indifferent, the existence of commitment devices will never affect the behavior of naifs in our model, since naïfs think they will always behave in the future according to their current preferences.

[15] By the present-bias effect, we mean the effect that the present bias has on the one-shot choice between doing it now versus doing it in some fixed future period. Note that for any one-shot choice, whether a person is sophisticated or naïve is irrelevant.

[16] Throughout this paper, "procrastination" means that an agent chooses to wait when her long-run self (i.e., a TC) would choose to do it, and "preproperation" means that an agent chooses to do it when her long-run self would choose to wait. We derived the word "perproperate" from the Latin root "praeproperum," which means "to do before the proper time." We later found this word in a few sufficiently unabridged dictionaries, with the definition we had intended.

[17] All propositions are stated with weak inequalities; but in each case, examples exist where the inequalities are strict. All proofs are in the Appendix.

Proposition 1 is as simple as it seems: Naïfs believe they will behave like TCs in the future but are more impatient now. Hence, the qualitative behavior of naïfs relative to TCs intuitively and solely reflects the present-bias effect.

The behavior of sophisticates is more complicated because there is a second effect influencing their behavior. The *sophistication effect* reflects that sophisticates are fully aware of any self-control problems they might have in the future, and this awareness can influence behavior now. The sophistication effect is captured in comparisons of sophisticates to naïfs. In our one-activity model, the sophistication effect is straightforward: Because sophisticates are (correctly) pessimistic that they will behave themselves in the future, they are more inclined than naïfs to do it now, irrespective of whether it is costs, rewards, or both that are immediate.

Proposition 2. For all cases $\tau_s \leq \tau_n$.

Even though sophisticates complete the activity before naïfs for both immediate costs and immediate rewards, the sophistication effect lends itself to different interpretations in these cases. For immediate costs, that sophisticates do it before naïfs reflects that sophistication helps mitigate the tendency to procrastinate, as discussed in example 1. For immediate rewards, that sophisticates do it before naïfs reflects that sophistication can exacerbate the tendency to preproperate, as discussed in example 2. These alternative interpretations will have important welfare implications, as we discuss in section 4.

Because sophisticates are influenced by the sophistication effect in addition to the present-bias effect, the qualitative behavior of sophisticates relative to TCs is complicated. In particular, it can be that sophisticates do not even exhibit the basic present-bias intuition. Consider the following scenario: Suppose you must write a paper this weekend, on Friday night, Saturday, or Sunday. You know the paper will be better if written on either Saturday or Sunday (when you have an entire day). However, it is a mid-November weekend with plenty of sports on TV—pro basketball on Friday night, college football on Saturday, and pro football on Sunday. You prefer watching pro football to college football, and prefer college football to pro basketball. Which sports event do you miss to write the paper? We can represent this scenario with the following example, where the activity to be done once is writing the paper and the costs correspond to the attractiveness of the sports event missed.

Example 3. Suppose costs are immediate, $T = 3$, and $\beta = 1/2$ for naïfs and sophisticates. Let $\mathbf{v} = \{12, 18, 18\}$ and $\mathbf{c} = \{3, 8, 13\}$.

Then $\tau_s = 1$ and $\tau_{tc} = 2$ (and $\tau_n = 3$).

TCs write the paper on Saturday because the marginal benefit of a better paper outweighs the marginal cost of giving up college football for pro basketball. Since the example involves immediate costs, the present-bias effect suggests that sophisticates should procrastinate. However, the sophistication effect leads sophisticates to write the paper on Friday night, before TCs. On Friday, sophisticates correctly predict that they will end up writing the paper on Sunday if they do not do it now. Hence, although sophisticates would prefer to write the paper on Saturday, they do it on Friday to prevent themselves from procrastinating until Sunday.

In example 3, sophisticates behave exactly opposite from what present-biased preferences would suggest, a result we will see again in sections 6 and 7. Of course, this is not always the case. Indeed, when rewards are immediate, sophisticates always preproperate because the sophistication effect exacerbates the self-control problem. Even so, situations like that in example 3 are not particularly pathological, and "preemptive overcontrol" is likely to arise in real-world environments (especially when choices are discrete). We highlight this result to emphasize the importance of sophistication effects. If you assume present-biased preferences and sophistication (as economists are prone to do), you must be careful to ask whether results are driven by present-biased preferences per se, or by present-biased preferences in conjunction with sophistication effects.

4. Welfare

Our emphasis in the previous section on qualitative behavioral comparisons among the three types of people masks what we feel may be a more important question about present-biased preferences: When does the taste for immediate gratification severely hurt a person? In this section, we examine the welfare implications of present-biased preferences with an eye towards this question. We show that even a small bias for the present can lead a person to suffer severe welfare losses, and characterize conditions when this can happen.

Welfare comparisons for people with time-inconsistent preferences are in principle problematic; the very premise of the model is that a person's preferences at different times disagree, so that a change in behavior may make some selves better off while making other selves worse off. The savings literature (e.g., Goldman 1979, 1980; Laibson 1994) often addresses this issue by defining a Pareto-efficiency criterion, asking when all period selves (weakly) prefer one strategy to another. If a strategy is Pareto superior to another, then it is clearly better. However, we feel this criterion is too strong: When applied to intertemporal choice, the Pareto criterion often refuses to rank two strategies even when one is much preferred by virtually all period selves, while the other is preferred by only one period self. Since present-biased preferences are often meant to capture self-control problems, where people pursue immediate gratification on a day-to-day basis, we feel the natural perspective in most situations is the "long-run perspective." (See Schelling [1984] for a thoughtful discussion of some of these issues.)[18]

To formalize the long-run perspective, we suppose there is a (fictitious) period 0 where the person has no decision to make and weights all future periods equally. We

[18] Indeed, Akerlof (1991) frames his discussion of procrastination in a way that emphasizes that a person's true preferences are her long-run preferences. Procrastination occurs in his model because costs incurred today are "salient"—a person experiences a cognitive illusion where costs incurred today loom larger than they are according to her true preferences.

can then denote a person's *long-run utility* from doing it in period τ by $U^0(\tau) \equiv v_\tau - c_\tau$. Our welfare analysis throughout this section will involve comparisons of long-run utilities. Even so, most of our welfare comparisons can be roughly conceived of as "Pareto comparisons," and we will note Pareto-efficiency "analogues" for our two main welfare results at the end of this section.

We begin with some brief qualitative comparisons of sophisticates and naïfs. The language in section 3 implied that sophistication is good when costs are immediate because it mitigates the tendency to procrastinate. Indeed, it is straightforward to show that when costs are immediate, sophisticates always do at least as well as naïfs [i.e., $U^0(\tau_s) \geq U^0(\tau_n)$]. Intuitively, since sophisticates never procrastinate in a period where naïfs do it, the only way their utilities can differ is when sophisticates preempt costly procrastination. When sophisticates choose to preempt costly procrastination, they do so despite their exaggerated aversion to incurring immediate costs, so this decision must also be preferred by the long-run self.

When rewards are immediate, on the other hand, the discussion in section 3 implied that sophistication is bad because it exacerbates the tendency to preproperate. More severe preproperation will often lead to lower long-run utility (as in example 2), but this is not necessarily the case. In particular, if there is a future period that is very tempting (i.e., it has a large reward) but very bad from a long-run perspective (i.e., it also has an even larger delayed cost), then more severe preproperation by sophisticates may in fact mean that sophisticates avoid this "temptation trap" while naifs do not. Hence, for immediate rewards we cannot say in general whether sophisticates or naifs are better off.

Rather than simple comparisons between sophisticates and naifs, however, our main focus for welfare analysis is the question of when a small bias for the present (i.e., β close to 1) can cause severe welfare losses. Since sophisticates, naifs, and TCs have identical long-run utility, we can measure the welfare loss from self-control problems by the deviation from TC long-run utility [i.e., $U^0(\tau_{tc}) - U^0(\tau_s)$ and $U^0(\tau_{tc}) - U^0(\tau_n)$].

We first note that if rewards and costs can be arbitrarily large, then a person with present-biased preferences can suffer arbitrarily severe welfare losses even from one-shot decisions. Suppose rewards are immediate, for instance, in which case a person with present-biased preferences is willing to grab a reward today for a delayed cost that is larger than the reward (by factor $1/\beta$). Even if β is very close to one, this decision can create an arbitrarily large welfare loss if the reward and cost are large enough.

We feel the more interesting case is when there is an upper bound on how large rewards and costs can be. In this case, the welfare loss from any individual bad decision will become very small as the self-control problem becomes small. But even if the welfare loss from any individual decision is small, severe welfare losses can still arise when self-control problems are compounded. To demonstrate this result, we suppose the upper bound on rewards and costs is $\overline{X}$. Then the welfare loss for both sophisticates and naifs cannot be larger than $2\overline{X}$.

Consider the case of immediate costs, where the self-control problem leads you to procrastinate. As in example 1, naïfs can compound self-control problems by making repeated decisions to procrastinate, each time believing they will do it next period. With each decision to procrastinate, they incur a small welfare loss, but the total welfare loss is the sum of these increments. No matter how small the individual welfare losses, naifs can suffer severe welfare losses if they procrastinate enough times. Sophisticates, in contrast, know exactly when they will do it if they wait, so delaying from period τ_{tc} to period τ_s is a single decision to procrastinate. Hence, for sophisticates small self-control problems cannot cause severe welfare losses. The following proposition formalizes these intuitions.

Proposition 3. Suppose costs are immediate, and consider all **v** and **c** such that $v_t \leq \overline{X}$ and $c_t \leq \overline{X}$ for all t:

(1) $\lim_{\beta \to 1} (\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_s)]) = 0$, and
(2) For any $\beta < 1$, $\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_n)] = 2\overline{X}$.

When rewards are immediate, however, and the self-control problem leads you to preproperate, we get the exact opposite result. For immediate rewards, naifs always believe that if they wait they will do it when TCs do it, so doing it in period τ_n as opposed to waiting until period τ_{tc} is a single decision to preproperate for naifs. Hence, for naïfs small self-control problems cannot cause severe welfare losses. But sophisticates can compound self-control problems because of an unwinding: In the end, sophisticates will preproperate; because they realize this, near the end they will preproperate; realizing this they preproperate a little sooner, etc. For each step of this unwinding, the welfare loss may be small, but the total welfare loss is the sum of multiple steps. As with naïfs and immediate costs, no matter how small the individual welfare losses, sophisticates can suffer severe welfare losses if the unraveling occurs over enough periods. These intuitions are formalized in proposition 4.

Proposition 4. Suppose rewards are immediate, and consider all **v** and **c** such that $v_t \leq \overline{X}$ and $c_t \leq \overline{X}$ for all t:

(1) $\lim_{\beta \to 1} (\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_n)]) = 0$, and
(2) For any $\beta < 1$, $\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_s)] = 2\overline{X}$.

As discussed at the beginning of this section, we feel that examining welfare losses in terms of long-run utility is the appropriate criterion to use when examining the welfare implications of present-biased preferences. Using this criterion, propositions 3 and 4 formalize when a small bias for the present can be very costly from a long-run perspective.[19] Even so, we note that there is also a less

[19] We feel that these limit results qualitatively capture very real differences in when moderately impatient sophisticates and naïfs can suffer severe welfare losses, but there are reasons to be cautious in interpreting them too literally. For instance, since "unwinding" drives severe preproperation for sophisticates, it seems natural to ask whether a small amount of uncertainty could reverse this tendency,

strong formalization using Pareto comparisons: If costs are immediate, sophisticates always choose a Pareto-optimal strategy while naifs may not; and if rewards are immediate, naifs always choose a Pareto-optimal strategy while sophisticates may not.

5. Smoking Guns

Many researchers studying time-inconsistent preferences have searched for empirical proof that people have such preferences. Efforts to indirectly prove time inconsistency have focused on the use of external "commitment devices" that limit future choice sets, because the use of such devices provides smoking guns that prove time consistency wrong. In this section, we show that smoking guns exist in our simple one-activity model, where no external commitment devices are available.

There are two properties that a person with time-consistent preferences will never violate. The first is "dominance": For intertemporal choice, one strategy dominates another if it yields in every period an instantaneous utility at least as large as the instantaneous utility from the other strategy, and strictly larger for some periods. In our model, one strategy is dominated by another if and only if the first strategy implies doing it at a cost with no reward while the second strategy implies doing it for a reward with no cost.[20]

Definition 5. A person obeys *dominance* if whenever there exists some period τ with $v_\tau > 0$ and $c_\tau = 0$ the person does not do it in any period τ' with $c_{\tau'} < 0$ and $v_{\tau'} = 0$.

much as Kreps et al. (1982) showed that a small amount of uncertainty can lead to extensive cooperation in the finitely repeated prisoner's dilemma. We suspect that there is something to this story, but the analogy is problematic on two fronts. First, although players may cooperate for most of a very long horizon, there is still a long duration at the end of the repeated prisoner's dilemma where players are unlikely to cooperate. Such an "endgame" could still create significant welfare losses. Second, in the Kreps et al. result a player's current behavior will signal something about her future behavior to other players. Since each "player"' in our game plays only once, the comparable signal is that a person in period t infers something about the propensity of her period-$(t + 1)$ self to wait from the fact that her period-$(t - 1)$ self waited, which requires that the period-t self does not know β. While we believe that such self-inference and self-signaling go on, there are many issues to be worked out to understand the strategic logic and psychological reality of such phenomena.

A comparable worry about our extreme results for naifs is that they will eventually learn that they have a tendency to procrastinate. Again, we think there is something to this intuition, but we suspect the issue is complicated. The issue of self-inference again arises. Further, people seem to have a powerful ability not to apply general lessons they understand well to specific situations. For instance, we are all familiar with the sensation of being simultaneously aware that we tend to be overoptimistic in completing projects, but *still* being overoptimistic regarding our current project. (See Kahneman and Lovallo [1993] for evidence on related issues.)

[20] E.g., consider a three-period example where $\mathbf{v} = (1, x, 0)$ and $\mathbf{c} = (0, y, 1)$. Then if costs are immediate, doing it in period 1 yields the stream of instantaneous utilities (0, 0, 0, 1) while doing it in period 3 yields the stream of instantaneous utilities $(-1, 0, 0, 0)$. Clearly the former dominates the latter.

The second property that a person with time-consistent behavior will never violate is independence of irrelevant alternatives—eliminating an option from the choice set that is not chosen should not change the person's choice from the remaining options.

Definition 6. For any $\mathbf{v} \equiv (v_1, v_2, \ldots, v_T)$ and $\mathbf{c} \equiv (c_1, c_2, \ldots, c_T)$, define

$$\mathbf{v}^{-t} \equiv (v_1, \ldots, v_{\tau-1}, v_{\tau+1}, \ldots, v_T) \text{ and}$$

$$\mathbf{c}^{-t} \equiv (c_1, \ldots, c_{\tau-1}, c_{\tau+1}, \ldots, c_T).$$

A person's behavior is *independent of irrelevant alternatives* if whenever she chooses period $\tau' \neq t$ when facing $\mathbf{v}$ and $\mathbf{c}$ she also chooses τ' when facing $\mathbf{v}^{-t}$ and $\mathbf{c}^{-t}$.

A time-consistent person will never violate dominance nor independence of irrelevant alternatives. These results hold for any time-consistent preferences, including time-consistent preferences that discount differently from period to period, and even time-consistent preferences that are not additively separable. Proposition 5 establishes that these results do not hold for people with present-biased preferences.

Proposition 5. For any β and δ such that $0 < \delta < 1$ and $0 < \beta < 1$, and for both sophistication and naïvete:

(1) There exists $(\mathbf{v}, \mathbf{c})$ and assumptions about immediacy such that a person with (β, δ)-preferences will violate dominance and

(2) There exists $(\mathbf{v}, \mathbf{c})$ and assumptions about immediacy such that a person with (β, δ)-preferences will violate independence of irrelevant alternatives.

To give some intuition for these results, we describe examples where each type violates dominance. The intuition for why each type violates independence of irrelevant alternatives is related. Sophisticates violate dominance when they choose a dominated early time to do an activity because they (correctly) worry that their future selves will not choose the dominating later time. For example, suppose rewards are immediate, $T = 3$, $\mathbf{v} = (0, 5, 1)$ and $\mathbf{c} = (1, 8, 0)$. Doing it in period 1 is clearly dominated by doing it in period 3. Even so, a sophisticate with $\beta = ½$ will complete the activity in period 1. She does so not because it is her most preferred period, but rather to avoid doing it in period 2. In period 1, the person prefers period 3 to period 1. Unfortunately, the period-2 self gets to choose between periods 2 and 3, and she will choose period 2.

Naïfs can violate dominance because of incorrect perceptions about future behavior. For example, suppose costs are immediate, $T = 3$, $\mathbf{v} = (1, 8, 0)$ and $\mathbf{c} = (0, 5, 1)$. Doing it in period 3 is dominated by doing it in period 1, and yet a naif with $\beta = ½$ will choose period 3. Even though in period 1 she prefers period 1 to period 3, she waits in period 1 incorrectly believing she will do it in period 2. Unfortunately, in period 2 she prefers waiting until period 3.[21]

Proposition *5* has important implications for the literature on smoking guns. First, proposition 5 implies that smoking guns need not involve the use of external commitment devices. Even simple behaviors can sometimes represent smoking guns. Furthermore, the literature on external commitment devices, provides smoking guns for sophisticates but not for naifs, since naifs would not pay to limit future choice sets. Proposition 5 implies that smoking guns exist for naifs as well. Finally, the intuitions above (and in the proof) suggest ways to design experiments attempting to find smoking guns, as well as the types of real world situations without external commitment devices where smoking guns might be found.

6. Multitasking

We now begin to explore how our results might carry over to more general settings. Consider a simple extension of our model where the activity must be performed more than once. The basic structure of the model is exactly as in section 2, but now the person must do the activity exactly $M \geq 1$ times, and she can do it at most once in any given period. We let $\tau'(M)$ denote the period in which a person completes the activity for the i^h time, and define $\Theta(M) \equiv \{\tau^1(M), \tau^2(M), \ldots, \tau^M(M)\}$. For each period τ in which the person does it, she receives reward θ_τ and incurs cost c_τ, and these can be experienced immediately or with some delay. Using the interpretations of immediate costs and immediate rewards from section 2, preferences take the following form.

1. *Immediate Costs*. Given $\Theta(M)$, the set of periods in which she does it, a person's intertemporal utility in period t is given by equation (1):

$$U^t(\Theta(M)) \equiv \begin{cases} -(1-\beta)c_t + \beta\left(\sum_{\tau\in\Theta(M)} \upsilon_t - \sum_{\tau\in\Theta(M)} c\right) & \text{if } t \in \Theta(M) \\ \beta\left(\sum_{\tau\in\Theta(M)} \upsilon_t - \sum_{\tau\in\Theta(M)} c_\tau\right) & \text{if } t \notin \Theta(M). \end{cases} \quad (1)$$

2. *Immediate Rewards*. Given $\Theta(M)$, the set of periods in which she does it, a person's intertemporal utility in period t is given by equation (2):

$$U^t(\Theta(M)) \equiv \begin{cases} (1-\beta)\upsilon_t + \beta\left(\sum_{\tau\in\Theta(M)} \upsilon_t - \sum_{\tau\in\Theta(M)} c_\tau\right) & \text{if } t \in \Theta(M) \\ \beta\left(\sum_{\tau\in\Theta(M)} \upsilon_t - \sum_{\tau\in\Theta(M)} c_\tau\right) & \text{if } t \notin \Theta(M). \end{cases} \quad (2)$$

[21] The proof of proposition 5 essentially involves generalizing these examples for all values of β and δ.

Given these preferences, we can define perception-perfect strategies analogously to Definitions 2, 3, and 4. We omit the formal definitions here. Let $\Theta_a(M) = \{\tau_n^t(M), \ldots, \tau_a^M(M)\}$ be the set of periods that an agent of type $a \in \{tc, s, n\}$ completes the activity according to her perception-perfect strategy. We begin by showing that the behavior of TCs and naifs in the multiactivity model is "normal" and intuitive.

Proposition 6. (1) For all cases and for any **v** and **c**, for each $M \in \{1, 2, \ldots, T - 1\}$; $\Theta_{tc}(M) \subset \Theta_{tc}(M + 1)$ and $\Theta_n(M) \subset \Theta_n(M + 1)$ and (2) If costs are immediate, then for all $i \in \{1, 2, \ldots, M\}$, $\tau_n^t(M) \geq \tau_{tc}^t(M)$, and if rewards are immediate, then for all $i \in \{1, 2, \ldots, M\}$, $\tau_n^s(M) \leq \tau_{tc}^t(M)$.

Part 1 of proposition 6 addresses how behavior depends on *M*: If TCs or naifs must do the activity an extra time, they do it in all periods they used to do it, and some additional period. If in any period they have *k* activities remaining, both TCs and naïfs do it now if and only if the current period is one of the *k* best remaining periods given their current preferences. Having more activities remaining, therefore, makes it more likely that they perform an activity now. Part 2 of proposition 6 states that the qualitative behavior of naifs relative to TCs in the multiactivity model is exactly analogous to that in the one-activity model. If costs are immediate, naïfs procrastinate: They are always behind TCs in terms of activities completed so far. If rewards are immediate, naifs preproperate: They are always ahead of TCs in terms of activities completed so far. Hence, the present-bias effect extends directly to the multiactivity setting; and again naifs exhibit the pure effects of present-biased preferences.

While the behavior of naifs in the multiactivity model is a straightforward and intuitive analogue of their behavior in the one-activity model, the effects of sophistication are significantly complicated. Consider the following example.

Example 4. Suppose rewards are immediate, $T = 3$, and $\beta = ½$ for naïfs and sophisticates. Let $\mathbf{v} = (6, 11, 21)$ and $\mathbf{c} = (0, 0, 0)$.

If $M = 1$, then $\tau_s = 1$, $\tau_n = 2$, and $\tau_{tc} = 3$.
If $M = 2$, then $\Theta_s(2) = \{2, 3\}$, $\Theta_n(2) = \{1, 2\}$, and $\Theta_{ic}(2) = \{2, 3\}$.

There are a couple of aspects of example 4 worth emphasizing. First, changing *M* dramatically changes the behavior of sophisticates: While sophisticates always preproperate when there is one activity, they do not preproperate here with two activities. Hence, the analogue to part 1 of proposition 6 does not hold for sophisticates. Sophisticates are always (looking for ways to influence their future behavior, and for $M > 1$ waiting can be a sort of "commitment device" to influence future behavior. If there is only one activity, there is no way to commit future selves not to preproperate. In example 4, when $M = 1$ the period-1 sophisticate does the activity because he (correctly) predicts that he will just do it in period 2 if he waits. If there is a second activity, however, a commitment device becomes available: Waiting now prevents you from doing the activity for the second time tomorrow: you can only do it for the first time tomorrow. Thus, forgoing the reward today makes you delay until period 3. When $M = 2$, the period-1 sophisticate knows he will do the second activity in period 2 if he does the first now, but he can force himself to do it in periods 2 and 3 if he waits now.

Example 4 also illustrates that the simple comparison of proposition 2—that for $M = 1$ sophisticates always do it before naïfs—does not extend to the multi-activity case. In example 4 with $M = 2$, sophisticates do it after naifs. The intuition behind proposition 2 was that sophisticates are correctly pessimistic about their utility from completing the activity in the future, and are therefore less willing to wait than naifs. But for $M > 1$ the relevant question is how pessimism affects the marginal utility of delaying one activity. As a result, there is no general result for the implications of sophistication versus naïvete. Example 4 shows for immediate rewards that sophistication can sometimes mitigate rather than exacerbate preproperation. Likewise, for immediate costs one can also find cases where sophistication exacerbates procrastination (and where sophisticates are worse off than naïfs). These examples illustrate that, in general environments, identifying when sophistication mitigates self-control problems and when it exacerbates them is more complicated than in the one-activity model. It is still true that sophisticates are more pessimistic than naifs about future behavior. But in more general environments, comparisons of sophisticates to naïfs depends on whether pessimism increases or decreases the marginal cost of current indulgence. As we discuss in section 7, in many contexts there are identifiable patterns as to how pessimism will affect incentives to behave oneself—but these patterns will not always correspond to the simple case of proposition 2.

We conclude this section by returning to a point made in section 3—that sophistication can lead a person to behave in ways that are seemingly contrary to having present-biased preferences. In section 3, we showed that sophisticates may do it before TCs even though costs are immediate. In the following example, sophisticates do things after TCs even though rewards are immediate.

Example 5. Suppose rewards are immediate, and $\beta = ½$ for naïfs and sophisticates. Let $\mathbf{v} = (12, 6, 11, 21)$ and $\mathbf{c} = (0, 0, 0, 0)$.

If $M = 2$, then $\Theta_{tc}(2) = \{1, 4\}$, $\Theta_n(2) = \{1, 3\}$, and $\Theta_s(2) = \{3, 4\}$.

In example 5, the situation beginning in period 2 is identical to example 4, and the intuition for why sophisticates do it later than TCs is related to the intuition of example 4. The period-1sophisticate knows that if he has one activity left in period 2, he will do it in period 2, while if he has two activities left in period 2, he will wait until periods 3 and 4. Hence, even though the period-1 sophisticate's most preferred periods for doing it are periods 1 and 4, he realizes he will not do it in period 4 if he does it in period 1. The choice for the period-1 sophisticate is between doing it in periods 1 and 2 versus doing it in periods 3 and 4. Of course situations like example 5 are somewhat special; but we do not feel they are so pathological that they will never occur in real-world environments (particularly for discrete choices).

7. Discussion and Conclusion

Many economic applications where present-biased preferences are clearly important cannot readily be put into the framework of this paper. Nonetheless, we feel our analysis provides some insight into such realms. In this section, we discuss some general lessons to take away from our analysis, and illustrate how these general lessons might play out in particular economic applications, such as savings and addiction.[22]

In our model, the behavior of naïfs intuitively and directly reflects their bias for the present. We suspect this simplicity in predicting the effects of naive self-control problems will hold in a broad array of economic models. Since consuming now yields immediate pay-offs whereas the increased future payoffs that saving allows is delayed, naifs will undersave in essentially any savings model; and since addictive activities involve yielding to some immediate desire today that has future costs naifs will overindulge in essentially any addiction model.

In contrast to naïfs, sophisticates in our model can behave in ways that seemingly contradict having present-biased preferences. We saw in section 3 that sophisticates may complete an unpleasant task before they would if they had no self-control problem, and in section 6 that they may consume tempting goods later than they would if they had no self-control problem. We suspect this complexity in predicting the effects of sophisticated self-control problems will also hold more generally. Sophistication effects that operate in addition to, and often in contradiction to, the present-bias effect can be quite significant. In the realm of saving, sophisticates can have a negative marginal propensity to consume over some ranges of income; and sophisticates can sometimes save more than TCs (i.e., they can behave exactly opposite from what a present bias would suggest).[23] In the realm of addiction,

[22] There has been much previous research on time inconsistency in savings models; see, for instance, Strotz (1956), Phelps and Pollak (1968), Pollak (1968), Thaler and Shefrin (1981), Shefrin and Thaler (1988, 1992), Laibson (1994, 1995, 1997), and Thaler (1994). Recently, economists have proposed models of "rational addiction" (Becker and Murphy, 1988; Becker et al., 1991, 1994). These models insightfully formalize the essence of (bad) addictive goods: Consuming more of the good today decreases overall utility but increases marginal utility for consumption of the same good tomorrow. However, these models a priori rule out the time-inconsistency and self-control issues modeled in this paper, and which many observers consider important in addiction.

[23] For simple examples of such behaviors, consider the following savings interpretation of a multi-activity model with $c = (0, 0, \ldots, 0)$: People have time-variant instantaneous utility functions, where in any period t the marginal utility of consuming the first dollar is v_t, and the marginal utility for any consumption beyond the first dollar is negligible. Then given wealth $M \in \{\$1, \$2, \ldots, \$T\}$ you must decide in which periods to consume. With this savings interpretation, sophisticates have a negative marginal propensity to consume in example 4: With wealth \$1, sophisticates consume \$1 in period 1, while with wealth \$2, sophisticates consume \$0 in period 1. And sophisticates save more than TCs in example 5: With wealth \$2, TCs consume \$1 in year 1 and save \$1 (which is consumed in year 4), while sophisticates consume \$0 in year 1 and save \$2 (which is consumed in years 3 and 4). Although examples 4 and 5 use rather special utility functions, it is relatively straightforward to find similar examples where utility functions are concave, increasing, and differentiable. We suspect, but have not proven, that sophisticates will never save more than TCs if utility functions are constant over time.

when it is optimal to consume an addictive product in moderation, sophisticates may not consume at all as a means of self-control—they know they will lose control if they try to consume in moderation. It is even possible to construct models where addictive goods are Giffen goods for sophisticates—non-addicts may buy more of a good in response to a permanent price increase, because high prices act as a sort of commitment device not to become addicted in the future.

People clearly have some degree of sophistication, and many sophistication effects—particularly attempts at self-control—seem very real. Other examples of sophistication effects seem perverse, however, and the corresponding behavior is likely to be somewhat rare. Hence, economists should be cautious when exploring present-biased preferences solely with the assumption of sophistication (which economists are prone to do since sophistication is closer to the standard economic assumptions). Because our analysis shows that sophistication effects can have large behavioral implications, and since people are clearly not completely sophisticated, researchers should be careful to clarify which results are driven by present-biased preferences per se, and which results arise from present-biased preferences in conjunction with sophistication effects.

We suspect one reason economists are so prone to assume sophistication in their models is the rule of thumb that less extreme departures from classical economic assumptions will lead to less extreme departures from classical predictions; hence, it is presumed that whatever novel predictions arise assuming sophistication will hold a fortiori assuming naivete. This rule of thumb does not apply here, of course, because many commitment strategies and other behaviors arise only because of sophistication. Moreover, our analysis also shows that even when sophistication does not affect the qualitative predictions, it does not always yield "milder" departures from conventional predictions: In many situations, being aware of self-control problems can exacerbate self-control problems.[24]

Indeed, another major theme of our analysis is to characterize the types of situations where sophistication mitigates versus exacerbates self-control problems. Extrapolating from our results, sophistication helps you when knowing about future misbehavior increases your perceived cost of current misbehavior, thereby encouraging you to behave yourself now. Sophistication hurts you when knowing about future misbehavior decreases the perceived cost of current misbehavior. In our one-activity model, this manifests itself in a simple fashion: When costs are immediate, you tend to procrastinate; if you are aware you will procrastinate in

[24] We have seen little discussion in the literature of how sophistication might affect the implications of self-control problems. Strotz (1956) and Akerlof (1991) discuss how sophistication might help improve behavior. We suspect their discussion reflects the prevalent intuition that sophistication can only help, and in fact have found no explicit discussion anywhere of how awareness of self-control problems might hurt. That sophistication can hurt you is, however, implicit in Pollak (1968). In the process of demonstrating a mathematical result, Pollak shows that sophisticates and naïfs behave the same for logarithmic utility. From this, it is straightforward to show that for utility functions more concave than the log utility function, sophisticates save more than naïfs (i.e., sophistication mitigates self-control problems), whereas for less concave utility functions, sophisticates save less than naïfs (i.e., sophistication exacerbates self-control problems).

the future, that makes you perceive it as more costly to procrastinate now. Hence, sophistication helps when costs are immediate. When rewards are immediate, you tend to preproperate; if you are aware you will preproperate in the future, that makes you perceive it as less costly to preproperate now. Hence, sophistication hurts when rewards are immediate.

In richer economic environments, whether sophistication helps or hurts will be more complicated. Nonetheless, our analysis suggests some simple conjectures. Consider, for example, the realm of addiction. Our analysis suggests sophistication might help when one wants to quit an addiction. A naïve person may repeatedly delay quitting smoking believing he will quit tomorrow; and proposition 4 suggests that this problem could lead to significant welfare losses. Sophistication should prevent this problem. In contrast, sophistication may hurt when a person is sure she will eventually get addicted, because this might lead to an unwinding logic along the lines of our example 2, by which she decides that since she will eventually succumb to temptation she might as well get addicted now.[25]

We conclude by reviewing two motivations for incorporating present-biased preferences into economic analysis. First, present-biased preferences may be useful in predicting behavior. There seem to be numerous applications where present-biased preferences can explain a prevalent behavior in a simple and plausible way, whereas post hoc and contrived explanations are required if one insists on interpreting phenomena through the prism of time-consistent preferences. For instance, Fischer (1997) observes that episodes of procrastination might be consistent with time consistency—but only if one assumes an absurd discount factor or implausibly low costs of delay. In contrast, present-biased preferences can explain the same episode of procrastination with a reasonable discount factor and a small bias for the present.[26]

But in many situations, present-biased preferences and time-consistent preferences both provide perfectly plausible explanations for behavior. Even so, a second motivation for incorporating present-biased preferences into economic analysis is that these two explanations can have vastly different welfare implications. For example, suppose a person becomes fat from eating large quantities of potato chips. She may do so because of a harmful self-control problem, or merely because the pleasure from eating potato chips outweighs the costs of being fat. Both hypotheses are reasonable explanations for the observed behavior: however, the two hypotheses have very different normative implications. The former says people buy too many potato chips at the prevailing price; the latter says they buy the right

[25] We believe it is likely that in most contexts—including addiction—sophistication will mitigate self-control problems rather than exacerbate them; but our analysis makes clear that there is no general principle guaranteeing this.

[26] O'Donoghue and Rabin (1999) show, in turn, that efforts to combat procrastination arising from present-biased preferences may help explain why incentive schemes involve deadlines that punish delays in completing a task much more harshly after some date than before that date—even when the true costs of delay are stationary. (Of course, it is likely there are plausible "time-consistent" explanations for the use of deadlines as well.)

amount. Because welfare analyses are often the main contribution economists can make, distinguishing between these two hypotheses is crucial. To further emphasize this point, consider the more policy-relevant example of an economic analysis of cigarette taxation that a priori assumes away self-control problems. This analysis may (or may not) yield a very accurate prediction of how cigarette taxes will affect consumption. But by ignoring self-control and related problems, it is likely to be either useless or very misleading as a guide to optimal cigarette-tax policy.

There are clearly many reasons to be cautious about welfare analyses that abandon rational-choice assumptions, and research ought to employ the most sophisticated methods available to carefully discern whether behaviors truly reflect harmful self-control problems. But the existence of present-biased preferences is overwhelmingly supported by psychological evidence, and strongly accords to common sense and conventional wisdom. And recall that our analysis in section 4 suggests that even relatively mild self-control problems can lead to significant welfare losses. Hence, even if the psychological evidence, common sense, and conventional wisdom are just a little right, and economists' habitual assumption of time consistency is just a little wrong, welfare economics ought be attentive to the role of self-control problems.

By analyzing the implications of present-biased preferences in a simple model, and positing some general lessons that will likely carry over to other contexts, we hope that our paper will add to other research in developing a tractable means for economists to investigate both the behavioral and welfare implications of present-biased preferences.

Appendix

Proof of Proposition 1

(1) We show that when costs are immediate, for any period if naïfs do it then TCs do it. Consider period t, and let $t' \equiv \max_{\tau > t}(v_\tau - c_\tau)$. Naïfs do it in period t only if $\beta v_t - c_t \geq \beta(v_{t'} - c_{t'})$, or $v_t - (1/\beta)c_t \geq v_{t'} - c_{t'}$; TCs do it in period t if $v_t - c_t \geq v_{t'} - c_{t'}$; and $v_t - c_t \geq v_t - (1/\beta)c_t$ for any $\beta \leq 1$. The result follows.

(2) We show that when rewards are immediate, for any period if TCs do it then naïfs do it. Consider period t, and let $t' \equiv \max_{\tau > t}(v_\tau - c_\tau)$. TCs do it in period t only if $v_t - c_t \geq v_{t'} - c_{t'}$; naifs do it in period t if $v_t - \beta c_t \geq \beta(v_{t'} - c_{t'})$, or $(1/\beta)v_t - c_t \geq v_{t'} - c_{t'}$; and $(1/\beta)v_t - c_t \geq v_t - c_t$ for any $\beta \leq 1$. The result follows.

Proof of Proposition 2

We show that for any period, if naïfs do it then sophisticates do it. Recall naïfs and sophisticates have identical preferences. The result follows directly because naïfs do it in period t only if $U^t(t) \geq U^t(\tau)$ for all $\tau > t$, while sophisticates do it in period t if $U^t(t) \geq U^t(\tau')$ for $\tau' = \min_{\tau>t}\{\tau \mid s^s_\tau = Y\}$.

Proof of Proposition 3

(1) We first argue that when costs are immediate, for any $t < t'$ such that $s_t^s = s_{t'}^s = Y$, $U^0(t) \geq U^0(t')$. This follows because for any t and $\tau' \equiv \min_{\tau > t}\{\tau \mid s_\tau^s = Y\}$, $s_t^s = Y$ only if $\beta v_t - c_t \geq \beta(v_{t'} - c_{t'})$, which implies $v_t - c_t \geq v_{t'} - c_{t'}$.

Now let $\bar{\tau} \equiv \min_{\tau > \tau_{tc}}\{\tau \mid s_\tau^s = Y\}$, so $\bar{\tau}$ is when sophisticates would do it if they waited in all $t \leq \tau_{tc}$. If $U^0(\tau_s) < U^0(\tau_{tc})$ then $s_{\tau_{tc}}^s = N$, so either $\tau_s = \bar{\tau}$ or $\tau_s < \tau_{tc}$. But using the result above, in either case $U^0(\tau_s) \geq U^0(\bar{\tau})$, which implies $U^0(\tau_{tc}) - U^0(\tau_s) \leq U^0(\tau_{tc}) - U^0(\bar{\tau})$. Given the definition of $\bar{\tau}$, $s_{\tau_{tc}}^s = N$ only if $\beta v_{\tau_{tc}} - c_{\tau_{tc}} - \beta U^0(\bar{\tau})$ or $-((1-\beta)/\beta)c_{\tau_{tc}} + U^0(\tau_{tc}) < U^0(\bar{\tau})$. Given the upper bound on costs $\overline{X}$, we must have $U^0(\tau_{tc}) - U^0(\tau_s) < ((1-\beta)/\beta)\overline{X}$. It is straightforward to show we can get arbitrarily close to this bound, so $\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_s)] = ((1-\beta)/\beta)\overline{X}$. Hence, $\lim_{\beta\to 1}(\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_s)]) = 0$.

(2) Fix $\beta < 1$. We will show that for any $\varepsilon \in (0, \overline{X})$ there exist reward/cost schedule combinations such that $U^0(\tau_{tc}) - U^0(\tau_n) = 2\overline{X} - \varepsilon$, from which the result follows. Choose $\gamma > 0$ such that $\beta + \gamma < 1$. Let i be the integer satisfying $(\varepsilon)/(\beta+\gamma)^i < \overline{X} \leq (\varepsilon)/(\beta+\gamma)^{i+1}$, and let j be the integer satisfying $\overline{X} - j((1-\beta)/(\beta+\gamma))\overline{X} > 0 \geq \overline{X} - (j+1)((1-\beta)/(\beta+\gamma))\overline{X}$. Consider the following reward and cost schedules where $T = i + j + 3$ is finite:

$$\mathbf{v} = (\overline{X}, \overline{X}, \ldots, \overline{X}, \overline{X} - ((1-\beta)/(\beta+\gamma))\overline{X}, \overline{X} - 2((1-\beta)/(\beta+\gamma))\overline{X}, \ldots, \overline{X} - j((1-\beta)/(\beta+\gamma))\overline{X}, 0)$$

$$\mathbf{c} = (\varepsilon, \varepsilon/(\beta+\gamma), \varepsilon/(\beta+\gamma)^2, \ldots, \varepsilon/(\beta+\gamma)^i, \overline{X}, \overline{X}, \ldots, \overline{X}).$$

Under $\mathbf{v}$ and $\mathbf{c}$, $\tau_{tc} = 1$ so $U^0(\tau_{tc}) = \overline{X} - \varepsilon$, and $\tau_n = T$ so $U^0(\tau_n) = -\bar{c}$. Hence, we have $U^0(\tau_{tc}) - U^0(\tau_n) = 2\overline{X} - \varepsilon$.

Proof of Proposition 4

(1) When rewards are immediate, by proposition 1 $\tau_n \leq \tau_{tc}$. For any $t < \tau_{tc}$, naïfs believe they will do it in period τ_{tc} if they wait. Hence, $v_{\tau_n} - \beta c_{\tau_n} \geq \beta U^0(t_{tc})$, which we can rewrite as $((1-\beta)/(\beta))\, v_{\tau_n} + U^0(\tau_n) \geq U^0(\tau_{tc})$. Given the upper bound on rewards $\overline{X}$, we have $U^0(\tau_{tc}) - U^0(\tau_n) \leq ((1-\beta)/\beta)\overline{X}$. Since the bound is easily achieved, $\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_n)] = ((1-\beta)/\beta)\overline{X}$, and $\lim_{\beta\to 1}(\sup_{(\mathbf{v},\mathbf{c})}[U^0(\tau_{tc}) - U^0(\tau_n)]) = 0$.

(2) Fix $\beta < 1$. We will show that for any $\varepsilon \in (0, \overline{X})$ there exist reward/cost schedule combinations such that $U^0(\tau_{tc}) - U^0(\tau_s) = 2\overline{X} - \varepsilon$, from which the result follows. Let i be the integer satisfying $(\varepsilon)/(\beta^i) < \overline{X} \leq (\varepsilon)/(\beta^{i+1})$, and let j be the integer satisfying $\overline{X} - j((1-\beta)/\beta)\overline{X} > 0 \geq \overline{X} - (j+1)((1-\beta)/\beta)\overline{X}$. Consider the following reward and cost schedules where $T = i + j + 3$ is finite:

$$\mathbf{v} = (\varepsilon, \varepsilon/(\beta), \varepsilon/(\beta^2), \ldots, \varepsilon/(\beta^i), \overline{X}, \overline{X}, \ldots, \overline{X})$$

$$\mathbf{c} = (\overline{X}, \overline{X}, \ldots, \overline{X}, \overline{X} - ((1-\beta)/\beta)\overline{X}, \overline{X} - 2((1-\beta)/\beta)\overline{X}, \ldots, \overline{X} - j((1-\beta)/\beta)\overline{X}, 0).$$

Under **v** and **c**, $\tau_{tc} = T$ so $U^0(\tau_{tc}) = \overline{X}$, and $\tau_s = 1$ so $U^0(\tau_s) = \varepsilon - \overline{X}$. Hence, we have $U^0(\tau_{tc}) - U^0(\tau_s) = 2\overline{X} - \varepsilon$.

Proof of Proposition 5

We prove each part by constructing examples.

(1) Suppose rewards are immediate, $T = 3$, $\mathbf{v} = (0, x, 1)$ and $\mathbf{c} = (1, y, 0)$. Sophisticates choose dominated strategy (Y, Y, Y) if $(x) - \beta\delta^2(y) \geq \beta\delta(1) - \beta\delta^2(0)$ and $0 - \beta\delta^3(1) \geq \beta\delta(x) - \beta\delta^3(y)$. We can rewrite these conditions as $\delta^2 y - \delta^2 \geq x \geq \beta\delta + \beta\delta^2 y$. If $y > (\beta + \delta)/(\delta(1 - \beta))$ then $\delta^2 y - \delta^2 > \beta\delta + \beta\delta^2 y$. Hence, for any β and δ there exists $y > (\beta + \delta)/(\delta(1 - \beta))$ and $x \in (\beta\delta + \beta\delta^2 y, \delta^2 y\beta - \delta^2)$, in which case $\mathbf{s}^s = (Y, Y, Y)$.

Suppose costs are immediate, $T = 3$, $\mathbf{y} = (1, x', 0)$ and $\mathbf{c} = (0, y', 1)$. Naïfs choose dominated strategy (N, N, Y) if $\beta\delta^3(1) - (0) < \beta\delta^3(x') - \beta\delta(y')$ and $\beta\delta^2(x') - (y') < \beta\delta^2(0) - \beta\delta(1)$. We can rewrite these conditions as $\delta^2 x' - \delta^2 > y' > \beta\delta + \beta\delta^2 x'$. If $x' > (\beta + \delta)/(\delta(1 - \beta))$ then $\delta^2 x' - \delta^2 > \beta\delta + \beta\delta^2 x'$. Hence, for any β and δ there exists $x' > (\beta + \delta)/(\delta(1 - \beta))$ and $y' \in (\beta\delta + \beta\delta^2 x', \delta^2 x' - \delta^2)$, in which case $\mathbf{s}^n = (N, N, Y)$.

(2) For any β and δ, choose $\phi \in (\sqrt{\beta}, 1)$, let $\mathbf{v} = (0, 0, 0)$ and $\mathbf{c} = (1, \phi/(\beta\delta), \phi^2/(\beta^2\delta^2))$, and suppose costs are immediate. Then sophisticates choose $\tau_s = 1$ when facing **v** and **c**, but $\tau_s = 2$ when facing $\mathbf{v}^{-\tau}$ and $\mathbf{c}^{-\tau}$, and this violates independence of irrelevant alternatives.

For any β and δ, choose $\phi \in (\sqrt{\beta}, 1)$, let $\mathbf{v} = (1, \phi/(\beta\delta), \phi^2(\beta^2\delta^2))$ and $\mathbf{c} = (0, 0, 0)$, and suppose rewards are immediate. Then naifs choose $\tau_n = 2$ when facing **v** and **c**, but $\tau_n = 1$ when facing $\mathbf{v}^{-\tau}$ and $\mathbf{c}^{-\tau}$, and this violates independence of irrelevant alternatives.

Proof of Proposition 6

(1) For both TCs and naifs, if they have k activities remaining in period t, then they do it in period t if and only if period t is one of the k best remaining periods given period-t preferences. Hence, for any $k' > k$, if TCs or naifs do it in period t with k activities remaining, then they do it in period t with k' activities remaining. Given this, the result is straightforward.

(2) We first show that for any t and k, when TCs and naifs each have k activities remaining in period t, then (i) for immediate costs if naifs do it in period t then TCs do it in period t; and (ii) for salient rewards if TCs do it in period t then naïfs do it in period t. Let t' be such that $v_t - c_t$ is the kth best $v_t - c_\tau$ for $\tau \in \{t + 1, t + 2, \ldots, T\}$. (i) follows because for immediate costs, naifs do it in period t only if $\beta v_t - c_t \geq \beta(v_{t'} - c_{t'})$, or $v_t - (1/\beta)c_{t'} \geq v_{t'} - c_{t'}$; TCs do it in period t if $v_t - c_t \geq v_{t'} - c_{t'}$; $v_t - c_t \geq v_t - (1/\beta)c_t$ for any $\beta \leq 1$. (ii) follows because for immediate rewards, TCs do it in period t only if $v_t - c_t \geq v_{t'} - c_{t'}$; naïfs do it in period t if $v_t - \beta c_t \geq \beta(v_{t'} - c_{t'})$, or $(1/\beta)v_t - c_t \geq v_{t'} - c_{t'}$; and $(1/\beta)\, v_t - c_t \geq v_t - c_t$ for any $\beta \leq 1$. The result then follows because (i) implies that for immediate costs

naïfs can never get ahead of TCs, and (ii) implies that for immediate rewards TCs can never get ahead of naïfs.

REFERENCES

Ainslie George. 1974. "Impulse Control in Pigeons." *Journal of the Experimental Analysis of Behavior*, 21(3): 485–89.

———. 1975. "Specious Reward: A Benavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin*, 82(4): 463–96.

———. 1987. "Self-Reported Tactics of Impulse Control." *International Journal of the Addictions*, 22(2): 167–79.

———. 1991. "Derivation of 'Rational' Economic Behavior from Hyperbolic Discount Curves." *American Economic Review* (*Papers and Proceedings*), 81(2): 334–40.

———. 1992. *Picoeconomics. The Strategic Interaction of Successive Motivational States within the Person*. New York: Cambridge University Press.

Ainslie, George, and Nick, Haslam. 1992a. "Self-Control." In *Choice over Time*, edited by George Loewenstein and Jon Elster. New York: Russell Sage.

———. 1992b. "Hyperbolic Discounting." In *Choice over Time*, edited by George Loewenstein and Jon Elster. New York: Russell Sage.

Ainslie, George, and Richard J. Herrnstein. 1981. "Preference Reversal and Delayed Reinforcement." *Animal Learning and Behavior*, 9(4): 476–82.

Akerlof, George A. 1991. "Procrastination and Obedience." *American Economic Review*, (*Papers and Proceedings*), 81(2): 1–19.

Becker, Gary S., Michael Grossman, and Kevin M. Murphy. 1991. "Rational Addiction and the Effect of Price on Consumption." *American Economic Review*, (*Papers and Proceedings*), 81(2): 237–41.

———. 1994. "An Empirical Analysis of Cigarette Addiction." *American Economic Review*, 84(3): 396–418.

Becker, Gary S., and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy*, 96(4): 675–701.

Chung, Shin-Ho, and Richard J. Herrnstein. 1967. "Choice and Delay of Reinforcement." *Journal of the Experimental Analysis of Behavior*, 10(1): 67–74.

Flscher, Carolyn. 1997. "Read This paper Even Later: Procrastination with Time-Inconsistent Preferences." Mimeo, University of Michigan.

Funder, David C., and Jack Block. 1989. "The Role of Ego-Control, Ego-Resiliency, and IQ in Delay of Gratification in Adolescence." *Journal of Personality and Social Psychology*, 57(6): 1041–50.

Glazer, Jacob, and Andrew Weiss. 1992. "Conflicting Preferences and Voluntary Restrictions on Choices." Mimeo, Tel Aviv University and Boston University.

Goldman, Steven M. 1979. "Intertemporally Inconsistent Preferences and the Rate of Consumption." *Econometrica*, 47(3): 621–26.

———. 1980. "Consistent Plans." *Review of Economic Studies*, 47(3): 533–37.

Hoch, Stephen J., and George Loewenstein. 1991. "Time-Inconsistent Preferences and Consumer Self-Control." *Journal of Consumer Research*, 17(4): 492–507.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1991. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives*, 5(1): 193–206.

Kahneman, Daniel, and Dan Lovalto. 1993. "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking." *Management Science*, 39(1): 17–31.
Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson. 1982. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory*, 27(2): 245–52.
Laibson, David. 1994. "Essays in Hyperbolic Discounting." Ph.D. diss. MIT.
———. 1995. "Hyperbolic Discount Functions, Undersaving, and Savings Policy." Mimeo, Harvard University.
———. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112(2): 443–77.
Loewenstein, George. 1992. "The Fall and Rise of Psychological Explanations in the Economics of Intertemporal Choice." In *Choice over Time*, edited by George Loewenstein and Jon Elster. New York: Russell Sage.
———. 1996. "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes*, 65(2): 272–92.
Loewenstein, George and Drazen Prelec. 1992. "Anomalies in Intertemporal Choice: Evidence and an Interpretation." *Quarterly Journal of Economics*, 107(2): 573–97.
O'Donoghue, Ted, and Matthew Rabin. 1996. "Doing It Now or Later." Math Center Discussion Paper No. 1172, Northwestern University.
———. 1997. "Incentives for Procrastinators." Math Center Discussion Paper No. 1181, Northwestern University, 1997; *Quarterly Journal of Economics*, 1999.
Peleg, Bezalel, and Menahem E. Yaart. 1973. "On the Existence of a Consistent Course of Action When Tastes Are Changing." *Review of Economic Studies*, 40(3): 391–401.
Phelps, Edmund S., and Robert A. Pollak. 1968. "On Second-Best National Saving and Game-Equilibrium Growth." *Review of Economic Studies*, 35(2): 185–99.
Pollak, Robert A. 1968. "Consistent Planning." *Review of Economic Studies*, 35(2): 201–8.
Prelec, Drazen. 1990. "Decreasing Impatience: Definition and Consequences." Working Paper No. 90–015, Harvard University.
Schelling, Thomas C. 1978. "Egonomics, or the Art of Self-Management." *American Economic Review* (*Papers and Proceedings*), 68(2): 290–94.
———. 1984. "Self-Command in Practice, in Policy, and in a Theory of Rational Choice." *American Economic Review*, (*Papers and Proceedings*), 74(2): 1–11.
———. 1992. "Self-Command: A New Discipline. In *Choice over Time*, edited by George Loewenstein and Jon Elster. New York: Russell Sage.
Shefrin, Hersh M., and Richard H. Thaler. 1988. "The Behavioral Life-Cycle Hypothesis." *Economic Inquiry*, 26(4): 609–43.
———. 1992. "Mental Accounting, Saving, and Self-Control." In *Choice over Time*, edited by George Loewenstein and Jon Elster. New York: Russell Sage.
Strotz, Robert H. 1956. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies*, 23(3): 165–80.
Thaler, Richard H. 1980. "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization*, 1(1): 39–60.
———. 1991. "Some Empirical Evidence on Dynamic Inconsistency." In *Quasi Rational Economics*, edited by Richard H. Thaler. New York: Russell Sage.
———. 1994. "Psychology and Savings Policies." *American Economic Review* (*Papers and Proceedings*), 84(2): 186–92.
Thaler, Richard H., and George Loewenstein. 1992. "Intertemporal Choice." In *The Winners' Curse: Paradoxes and Anomalies of Economic Life*, edited by Richard H. Thaler. New York: Free Press.

Thaler, Richard H., and Hersh M. Shefrin. 1981. "An Economic Theory of Self-Control." *Journal of Political Economy*, 89(2): 392–406.

Watterson, Bill. 1993. *The Days Are Just Packed: A Calvin and Hobbes Collection*. Kansas City, MO: Andrews and McMeel.

Wertenbroch, Klaus. 1993. "Consumer Self-Control: How Much Vice and Virtue Do We Buy?" Mimeo, University of Chicago.

Yaari, Menahem E. 1977. "Consistent Utilization of an Exhaustible Resource, or, How to Eat an Appetite-Arousing Cake." Center for Research in Mathematical Economics and Game Theory Research Memorandum No. 26.

CHAPTER 8

Fairness as a Constraint on Profit Seeking: Entitlements in the Market

DANIEL KAHNEMAN, JACK L. KNETSCH, AND RICHARD H. THALER

JUST AS it is often useful to neglect friction in elementary mechanics, there may be good reasons to assume that firms seek their maximal profit as if they were subject only to legal and budgetary constraints. However, the patterns of sluggish or incomplete adjustment often observed in markets suggest that some additional constraints are operative. Several authors have used a notion of fairness to explain why many employers do not cut wages during periods of high unemployment (Akerlof 1979; Solow 1980). Okun (1981) went further in arguing that fairness also alters the outcomes in what he called customer markets—characterized by suppliers who are perceived as making their own pricing decisions, have some monopoly power (if only because search is costly), and often have repeat business with their clientele. Like labor markets, customer markets also sometimes fail to clear:

> [F]irms in the sports and entertainment industries offer their customers tickets at standard prices for events that clearly generate excess demand. Popular new models of automobiles may have waiting lists that extend for months. Similarly, manufacturers in a number of industries operate with backlogs in booms and allocate shipments when they obviously could raise prices and reduce the queue. (p. 170)

Okun explained these observations by the hostile reaction of customers to price increases that are not justified by increased costs and are therefore viewed as unfair. He also noted that customers appear willing to accept "fair" price increases even when demand is slack, and commented that "in practice, observed pricing behavior is vast distance from do it yourself auctioneering" (p. 170).

The argument used by these authors to account for apparent deviations from the simple model of a profit-maximizing firm is that fair behavior is instrumental to the maximization of long-run profits. In Okun's model, customers who suspect that a supplier treats them unfairly are likely to start searching for alternatives;

The research was carried out when Kahneman was at the University of British Columbia. It was supported by the Department of Fisheries and Oceans Canada. Kahneman and Thaler were also supported by the U.S. Office of Naval Research and the Alfred P. Sloan Foundation, respectively. Conversations with J. Brander, R. Frank, and A. Tversky were very helpful.

Akerlof (1980, 1982) suggested that firms invest in their reputation to produce goodwill among their customers and high morale among their employees; and Arrow argued that trusted suppliers may be able to operate in markets that are otherwise devastated by the lemons problem (Akerlof 1970; Arrow 1973). In these approaches, the rules of fairness define the terms of an enforceable implicit contract: Firms that behave unfairly are punished in the long run. A more radical assumption is that some firms apply fair policies even in situations that preclude enforcement—this is the view of the lay public, as shown in a later section of this chapter.

If considerations of fairness do restrict the actions of profit-seeking firms, economic models might be enriched by a more detailed analysis of this constraint. Specifically, the rules that govern public perceptions of fairness should identify situations in which some firms will fail to exploit apparent opportunities to increase their profits. Near-rationality theory (Akerlof and Yellen 1985) suggests that such failures to maximize by a significant number of firms in a market can have large aggregate effects even in the presence of other firms that seek to take advantage of all available opportunities. Rules of fairness can also have significant economic effects through the medium of regulation. Indeed, Edward Zajac (forthcoming) has inferred general rules of fairness from public reactions to the behavior of regulated utilities.

The present research uses household surveys of public opinions to infer rules of fairness for conduct in the market from evaluations of particular actions by hypothetical firms.[1] The study has two main objectives: (1) to identify community standards of fairness that apply to price, rent, and wage setting by firms in varied circumstances; and (2) to consider the possible implications or the rules of fairness for market outcomes.

The study was concerned with scenarios in which a *firm* (merchant, landlord, or employer) makes a pricing or wage-setting decision that affects the outcomes of one or more *transactors* (customers, tenants, or employees). The scenario was read to the participants, who evaluated the fairness of the action as in the following example:

Question 1. A hardware store has been selling snow shovels for $15. The morning after a large snowstorm, the store raises the price to $20. Please rate this action as:

Completely Fair	Acceptable
Unfair	Very Unfair

The two favorable and the two unfavorable categories are grouped in this report to indicate the proportions of respondents who judged the action acceptable or unfair. In this example, 82 percent of respondents ($N = 107$) considered it unfair

[1] Data were collected between May 1984 and July 1985 in telephone surveys of randomly selected residents of two Canadian metropolitan areas: Toronto and Vancouver. Equal numbers of adult female and male respondents were interviewed for about ten minutes in calls made during evening hours. No more than five questions concerned with fairness were included in any interview, and contrasting questions that were to be compared were never put to the same respondents.

for the hardware store to take advantage of the short-run increase in demand associated with a blizzard.

The approach of the present study is purely descriptive. Normative status is not claimed for the generalizations that are described as "rules of fairness," and the phrase "it is fair" is simply an abbreviation for "a substantial majority of the population studied thinks it fair." The chapter considers in turn three determinants of fairness judgments: the reference transaction, the outcomes to the firm and to the transactors, and the occasion for the action of the firm. The final sections are concerned with the enforcement of fairness and with economic phenomena that the rules of fairness may help explain.

1. Reference Transactions

A central concept in analyzing the fairness of actions in which a firm sets the terms of future exchanges is the *reference transaction*, a relevant precedent that is characterized by a reference price or wage, and by a positive reference profit to the firm. The treatment is restricted to cases in which the fairness of the reference transaction is not itself in question.

The main findings of this research can be summarized by a principle of *dual entitlement*, which governs community standards of fairness: Transactors have an entitlement to the terms of the reference transaction and firms are entitled to their reference profit. A firm is not allowed to increase its profits by arbitrarily violating the entitlement of its transactors to the reference price, rent or wage (Bazerman 1985; Zajac, forthcoming). When the reference profit of a firm is threatened, however, it may set new terms that protect its profit at transactors' expense.

Market prices, posted prices, and the history of previous transactions between a firm and a transactor can serve as reference transactions. When there is a history of similar transactions between firm and transactor, the most recent price, wage, or rent will be adopted for reference unless the terms of the previous transaction were explicitly temporary. For new transactions, prevailing competitive prices or wages provide the natural reference. The role of prior history in wage transactions is illustrated by the following pair of questions:

Question 2A. A small photocopying shop has one employee who has worked in the shop for six months and earns $9 per hour. Business continues to be satisfactory, but a factory in the area has closed and unemployment has increased. Other small shops have now hired reliable workers at $7 an hour to perform jobs similar to those done by the photocopy shop employee. The owner of the photocopying shop reduces the employee's wage to $7.

($N = 98$) Acceptable 17% Unfair 83%

Question 2B. A small photocopying shop has one employee [as in Question 2A]. The current employee leaves, and the owner decides to pay a replacement $7 an hour.

($N = 125$) Acceptable 73% Unfair 27%

The current wage of an employee serves as reference for evaluating the fairness of future adjustments of that employee's wage—but not necessarily for evaluating the fairness of the wage paid to a replacement. The new worker does not have an entitlement to the former worker's wage rate. As the following question shows, the entitlement of an employee to a reference wage does not carry over to a new labor transaction, even with the same employer:

Question 3. A house painter employs two assistants and pays them $9 per hour. The painter decides to quit house painting and go into the business of providing landscape services, where the going wage is lower. He reduces the workers' wages to $7 per hour for the landscaping work.

(N = 94) Acceptable 63% Unfair 37%

Note that the same reduction in wages that is judged acceptable by most respondents in question 3 was judged unfair by 83 percent of the respondents to question 2A.

Parallel results were obtained in questions concerning residential tenancy. As in the case of wages, many respondents apply different rules to a new tenant and to a tenant renewing a lease. A rent increase that is judged fair for a new lease may be unfair for a renewal. However, the circumstances under which the rules of fairness require landlords to bear such opportunity costs are narrowly defined. Few respondents consider it unfair for the landlord to sell the accommodation to another landlord who intends to raise the rents of sitting tenants, and even fewer believe that a landlord should make price concessions in selling an accommodation to its occupant.

The relevant reference transaction is not always unique. Disagreements about fairness are most likely to arise when alternative reference transactions can be invoked, each leading to a different assessment of the participants' outcomes. Agreement on general principles of fairness therefore does not preclude disputes about specific cases (see also Zajac, forthcoming). When competitors change their price or wage, for example, the current terms set by the firm and the new terms set by competitors define alternative reference transactions. Some people will consider it unfair for a firm not to raise its wages when competitors are increasing theirs. On the other hand, price increases that are not justified by increasing costs are judged less objectionable when competitors have led the way.

It should perhaps be emphasized that the reference transaction provides a basis for fairness judgments because it is normal, not necessarily because it is just. Psychological studies of adaptation suggest that any stable state of affairs tends to become accepted eventually, at least in the sense that alternatives to it no longer readily come to mind. Terms of exchange that are initially seen as unfair may in time acquire the status of a reference transaction. Thus, the gap between the behavior that people consider fair and the behavior that they expect in the market-place tends to be rather small. This was confirmed in several scenarios, where different samples of respondents answered the two questions: "What does fairness require?" and "What do you think the firm would do?" The similarity of the answers suggests that people expect a substantial level of conformity to community standards—and also that they adapt their views of fairness to the norms of actual behavior.

2. The Coding of Outcomes

It is a commonplace that the fairness of an action depends in large part on the signs of its outcomes for the agent and for the individuals affected by it. The cardinal rule of fair behavior is surely that one person should not achieve a gain by simply imposing an equivalent loss on another.

In the present framework, the outcomes to the firm and to its transactors are defined as gains and losses in relation to the reference transaction. The transactor's outcome is simply the difference between the new terms set by the firm and the reference price, rent, or wage. The outcome to the firm is evaluated with respect to the reference profit, and incorporates the effect of exogenous shocks (for example, changes in wholesale prices) which alter the profit of the firm on a transaction at the reference terms. According to these definitions, the outcomes in the snow shovel example of question 1 were a $5 gain to the firm and a $5 loss to the representative customer. However, had the same price increase been induced by a $5 increase in the wholesale price of snow shovels, the outcome to the firm would have been nil.

The issue of how to define relevant outcomes takes a similar form in studies of individuals' preferences and of judgments of fairness. In both domains, a descriptive analysis of people's judgments and choices involves rules of *naïve accounting* that diverge in major ways from the standards of rationality assumed in economic analysis. People commonly evaluate outcomes as gains or losses relative to a neutral reference point rather than as endstates (Kahneman and Tversky 1979). In violation of normative standards, they are more sensitive to out-of-pocket costs than to opportunity costs and more sensitive to losses than to foregone gains (Kahneman and Tversky 1984; Thaler 1980). These characteristics of evaluation make preferences vulnerable to framing effects, in which inconsequential variations in the presentation of a choice problem affect the decision (Tversky and Kahneman 1986).

The entitlements of firms and transactors induce similar asymmetries between gains and losses in fairness judgments. An action by a firm is more likely to be judged unfair if it causes a loss to its transactor than if it cancels or reduces a possible gain. Similarly, an action by a firm is more likely to be judged unfair if it achieves a gain to the firm than if it averts a loss. Different standards are applied to actions that are elicited by the threat of losses or by an opportunity to improve on a positive reference profit—a psychologically important distinction which is usually not represented in economic analysis.

Judgments of fairness are also susceptible to framing effects, in which form appears to overwhelm substance. One of these framing effects will be recognized as the money illusion, illustrated in the following questions:

Question 4A. A company is making a small profit. It is located in a community experiencing a recession with substantial unemployment but no inflation. There are many workers anxious to work at the company. The company decides to decrease wages and salaries 7% this year.

($N = 125$) Acceptable 38% Unfair 62%

Question 4B. With substantial unemployment and inflation of 12% . . . the company decides to increase salaries only 5% this year.

($N = 129$) Acceptable 78% Unfair 22%

Although the real income change is approximately the same in the two problems, the judgments of fairness are strikingly different. A wage cut is coded as a loss and consequently judged unfair. A nominal raise which does not compensate for inflation is more acceptable because it is coded as a gain to the employee, relative to the reference wage.

Analyses of individual choice suggest that the disutility associated with an outcome that is coded as a loss may be greater than the disutility of the same objective outcome when coded as the elimination of a gain. Thus, there may be less resistance to the cancellation of a discount or bonus than to an equivalent price increase or wage cut. As illustrated by the following questions, the same rule applies as well to fairness judgments.

Question 5A. A shortage has developed for a popular model of automobile, and customers must now wait two months for delivery. A dealer has been selling these cars at list price. Now the dealer prices this model at $200 above list price.

($N = 130$) Acceptable 29% Unfair 71%

Question 5B. A dealer has been selling these cars at a discount of $200 below list price. Now the dealer sells this model only at list price.

($N = 123$) Acceptable 58% Unfair 42%

The significant difference between the responses to questions 5A and 5B (*chi*-squared = 20.91) indicates that the $200 price increase is not treated identically in the two problems. In question 5A the increase is clearly coded as a loss relative to the unambiguous reference provided by the list price. In question 5B the reference price is ambiguous, and the change can be coded either as a loss (if the reference price is the discounted price), or as the elimination of a gain (if the reference price is the list price). The relative leniency of judgments in question 5B suggests that at least some respondents adopted the latter frame. The following questions illustrate the same effect in the case of wages:

Question 6A. A small company employs several people. The workers' incomes have been about average for the community. In recent months, business for the company has not increased as it had before. The owners reduce the workers' wages by 10 percent for the next year.

($N = 100$) Acceptable 39% Unfair 61%

Question 6B. A small company employs several people. The workers have been receiving a 10 percent annual bonus each year and their total incomes have been about average for the community. In recent months, business for the company has not increased as it had before. The owners eliminate the workers' bonus for the year.

($N = 98$) Acceptable 80% Unfair 20%

3. Occasions for Pricing Decisions

This section examines the rules of fairness that apply to three classes of occasions in which a firm may reconsider the terms that it sets for exchanges. (1) *Profit reductions*, for example, by rising costs or decreased demand for the product of the firm. (2) *Profit increases*, for example, by efficiency gains or reduced costs. (3) *Increases in market power*, for example, by temporary excess demand for goods, accommodations or jobs.

Protecting Profit

A random sample of adults contains many more customers, tenants, and employees than merchants, landlords, or employers. Nevertheless, most participants in the surveys clearly consider the firm to be entitled to its reference profit: They would allow a firm threatened by a reduction of its profit below a positive reference level to pass on the entire loss to its transactors, without compromising or sharing the pain. By large majorities, respondents endorsed the fairness of passing on increases in wholesale costs, in operating costs, and in the costs associated with a rental accommodation. The following two questions illustrate the range of situations to which this rule was found to apply.

Question 7. Suppose that, due to a transportation mixup, there is a local shortage of lettuce and the wholesale price has increased. A local grocer has bought the usual quantity of lettuce at a price that is 30 cents per head higher than normal. The grocer raises the price of lettuce to customers by 30 cents per head.

($N = 101$) Acceptable 79% Unfair 21%

Question 8. A landlord owns and rents out a single small house to a tenant who is living on a fixed income. A higher rent would mean the tenant would have to move. Other small rental houses are available. The landlord's costs have increased substantially over the past year and the landlord raises the rent to cover the cost increases when the tenant's lease is due for renewal.

($N = 151$) Acceptable 75% Unfair 25%

The answers to the last question, in particular, indicate that it is acceptable for firms to protect themselves from losses even when their transactors suffer substantial inconvenience as a result. The rules of fairness that yield such judgments do not correspond to norms of charity and do not reflect distributional concerns.

The attitude that permits the firm to protect a positive reference profit at the transactors' expense applies to employers as well as to merchants and landlords. When the profit of the employer in the labor transaction falls below the reference level, reductions of even nominal wages become acceptable. The next questions illustrate the strong effect of this variable.

Question 9A. A small company employs several workers and has been paying them average wages. There is severe unemployment in the area and the company could easily replace its current employees with good workers at a lower wage. The company has been making money. The owners reduce the current workers' wages by 5 percent.

(N = 195) Acceptable 23% Unfair 77%

Question 9B. The company has been losing money. The owners reduce the current workers' wages by 5 percent.

(N = 195) Acceptable 68% Unfair 32%

The effect of firm profitability was studied in greater detail in the context of a scenario in which Mr. Green, a gardener who employs two workers at $7 an hour, learns that other equally competent workers are willing to do the same work for $6 an hour. Some respondents were told that Mr. Green's business was doing well, others were told that it was doing poorly. The questions, presented in open format, required respondents to state "what is fair for Mr. Green to do in this situation," or "what is your best guess about what Mr. Green would do." The information about the current state of the business had a large effect. Replacing the employees or bargaining with them to achieve a lower wage was mentioned as fair by 67 percent of respondents when business was said to be poor, but only by 25 percent of respondents when business was good. The proportion guessing that Mr. Green would try to reduce his labor costs was 75 percent when he was said to be doing poorly, and 49 percent when he was said to be doing well. The differences were statistically reliable in both cases.

A firm is only allowed to protect itself at the transactor's expense against losses that pertain directly to the transaction at hand. Thus, it is unfair for a landlord to raise the rent on an accommodation to make up for the loss of another source of income. On the other hand, 62 percent of the respondents considered it acceptable for a landlord to charge a higher rent for apartments in one of two otherwise identical buildings, because a more costly foundation had been required in the construction of that building.

The assignment of costs to specific goods explains why it is generally unfair to raise the price of old stock when the price of new stock increases:

Question 10. A grocery store has several months supply of peanut butter in stock which it has on the shelves and in the storeroom. The owner hears that the wholesale price of peanut butter has increased and immediately raises the price on the current stock of peanut butter.

(N = 147) Acceptable 21% Unfair 79%

The principles of naive accounting apparently include a FIFO method of inventory cost allocation.

The Allocation of Gains

The data of the preceding section could be interpreted as evidence for a cost-plus rule of fair pricing, in which the supplier is expected to act as a broker in passing on marked-up costs (Okun). A critical test of this possible rule arises when the supplier's costs diminish: A strict cost-plus rule would require prices to come down accordingly. In contrast, a dual-entitlement view suggests that the firm is only prohibited from increasing its profit by causing a loss to its transactors. Increasing profits by retaining cost reductions does not violate the transactors' entitlement and may therefore be acceptable.

The results of our previous study (1986) indicated that community standards of fairness do not in fact restrict firms to the reference profit when their costs diminish, as a cost-plus rule would require. The questions used in these surveys presented a scenario of a monopolist supplier of a particular kind of table, who faces a \$20 reduction of costs on tables that have been selling for \$150. The respondents were asked to indicate whether "fairness requires" the supplier to lower the price, and if so, by how much. About one-half of the survey respondents felt that it was acceptable for the supplier to retain the entire benefit, and less than one-third would require the supplier to reduce the price by \$20, as a cost-plus rule dictates. Further, and somewhat surprisingly, judgments of fairness did not reliably discriminate between primary producers and middlemen, or between savings due to lower input prices and to improved efficiency.

The conclusion that the rules of fairness permit the seller to keep part or all of any cost reduction was confirmed with the simpler method employed in the present study.

Question 11A. A small factory produces tables and sells all that it can make at \$200 each. Because of changes in the price of materials, the cost of making each table has recently decreased by \$40. The factory reduces its price for the tables by \$20.

(N = 102) Acceptable 79% Unfair 21%

Question 11B. The cost of making each table has recently decreased by \$20. The factory does not change its price for the tables.

(N = 100) Acceptable 53% Unfair 47%

The even division of opinions on question 11B confirms the observations of the previous study. In conjunction with the results of the previous section, the findings support a dual-entitlement view: the rules of fairness permit a firm not to share in the losses that it imposes on its transactors, without imposing on it an unequivocal duty to share its gains with them.

Exploitation of Increased Market Power

The market power of a firm reflects the advantage to the transactor of the exchange which the firm offers, compared to the transactor's second-best alternative. For example, a blizzard increases the surplus associated with the purchase of a snow shovel at the regular price, compared to the alternatives of buying elsewhere or doing without a shovel. The respondents consider it unfair for the hardware store to capture any part of the increased surplus, because such an action would violate the customer's entitlement to the reference price. Similarly, it is unfair for a firm to exploit an excess in the supply of labor to cut wages (question 2A), because this would violate the entitlement of employees to their reference wage.

As shown by the following routine example, the opposition to exploitation of shortages is not restricted to such extreme circumstances:

Question 12. A severe shortage of Red Delicious apples has developed in a community and none of the grocery stores or produce markets have any of this type of apple on their shelves. Other varieties of apples are plentiful in all of the stores. One grocer receives a single shipment of Red Delicious apples at the regular wholesale cost and raises the retail price of these Red Delicious apples by 25% over the regular price.

(N = 102) Acceptable 37% Unfair 63%

Raising prices in response to a shortage is unfair even when close substitutes are readily available. A similar aversion to price rationing held as well for luxury items. For example, a majority of respondents thought it unfair for a popular restaurant to impose a $5 surcharge for Saturday night reservations.

Conventional economic analyses assume as a matter of course that excess demand for a good creates an opportunity for suppliers to raise prices, and that such increases will indeed occur. The profit-seeking adjustments that clear the market are in this view as natural as water finding its level—and as ethically neutral. The lay public does not share this indifference. Community standards of fairness effectively require the firm to absorb an opportunity cost in the presence of excess demand, by charging less than the clearing price or paying more than the clearing wage.

As might be expected from this analysis, it is unfair for a firm to take advantage of an increase in its monopoly power. Respondents were nearly unanimous in condemning a store that raises prices when its sole competitor in a community is temporarily forced to close. As shown in the next question, even a rather mild exploitation of monopoly power is considered unfair.

Question 13. A grocery chain has stores in many communities. Most of them face competition from other groceries. In one community the chain has no competition. Although its costs and volume of sales are the same there as elsewhere, the chain sets prices that average 5 percent higher than in other communities.

(N = 101) Acceptable 24% Unfair 76%

Responses to this and two additional versions of this question specifying average price increases of 10 and 15 percent did not differ significantly. The respondents clearly viewed such pricing practices as unfair, but were insensitive to the extent of the unwarranted increase.

A monopolist might attempt to increase profits by charging different customers as much as they are willing to pay. In conventional theory, the constraints that prevent a monopolist from using perfect price discrimination to capture all the consumers' surplus are asymmetric information and difficulties in preventing resale. The survey results suggest the addition of a further restraint: some forms of price discrimination are outrageous.

Question 14. A landlord rents out a small house. When the lease is due for renewal, the landlord learns that the tenant has taken a job very close to the house and is therefore unlikely to move. The landlord raises the rent \$40 per month more than he was planning to do.

($N = 157$) Acceptable 9% Unfair 91%

The near unanimity of responses to this and similar questions indicates that an action that deliberately exploits the special dependence of a particular individual is exceptionally offensive.

The introduction of an explicit auction to allocate scarce goods or jobs would also enable the firm to gain at the expense of its transactors, and is consequently judged unfair.

Question 15. A store has been sold out of the popular Cabbage Patch dolls for a month. A week before Christmas a single doll is discovered in a storeroom. The managers know that many customers would like to buy the doll. They announce over the store's public address system that the doll will be sold by auction to the customer who offers to pay the most.

($N = 101$) Acceptable 26% Unfair 74%

Question 16. A business in a community with high unemployment needs to hire a new computer operator. Four candidates are judged to be completely qualified for the job. The manager asks the candidates to state the lowest salary they would be willing to accept, and then hires the one who demands the lowest salary.

($N = 154$) Acceptable 36% Unfair 64%

The auction is opposed in both cases, presumably because the competition among potential buyers or employees benefits the firm. The opposition can in some cases be mitigated by eliminating this benefit. For example, a sentence added to question 15, indicating that "the proceeds will go to UNICEF" reduced the negative judgments of the doll auction from 74 to 21 percent.

The strong aversion to price rationing in these examples clearly does not extend to all uses of auctions. The individual who sells securities at twice the price paid for them a month ago is an object of admiration and envy—and is certainly not thought to be gouging. Why is it fair to sell a painting or a house at the market-clearing price, but not an apple, dinner reservation, job, or football game ticket? The rule of acceptability appears to be this: Goods for which an active resale market exists, and especially goods that serve as a store of value, can be sold freely by auction or other mechanisms allowing the seller to capture the maximum price.

When resale is a realistic possibility, which is not the case for most consumer goods, the potential resale price reflects the higher value of the asset and the purchaser is therefore not perceived as sustaining a loss.

4. Enforcement

Several considerations may deter a firm from violating community standards of fairness. First, a history or reputation of unfair dealing may induce potential transactors to take their business elsewhere, because of the element of trust that is present in many transactions. Second, transactors may avoid exchanges with offending firms at some cost to themselves, even when trust is not an issue. Finally, the individuals who make decisions on behalf of firms may have a preference for acting fairly. The role of reputation effects is widely recognized. This section presents some indications that a willingness to resist and to punish unfairness and an intrinsic motivation to be fair could also contribute to fair behavior in the marketplace.

A willingness to pay to resist and to punish unfairness has been demonstrated in incentive compatible laboratory experiments. In the ultimatum game devised by Werner Guth, Rolf Schmittberger, and Bernd Schwarze (1982), the participants are designated as allocators or recipients. Each allocator anonymously proposes a division of a fixed amount of money between himself (herself) and a recipient. The recipient either accepts the offer or rejects it, in which case both players get nothing. The standard game theoretic solution is for the allocator to make a token offer and for the recipient to accept it, but Guth et al. observed that many allocators offer an equal division and that recipients sometimes turn down positive offers. In our more detailed study of resistance to unfairness (1986), recipients were asked to indicate in advance how they wished to respond to a range of possible allocations: A majority of participants were willing to forsake $2 rather than accept an unfair allocation of $10.

Willingness to punish unfair actors was observed in another experiment, in which subjects were given the opportunity to share a sum of money evenly with one of two anonymous strangers, identified only by the allocation they had proposed to someone else in a previous round. About three-quarters of the undergraduate participants in this experiment elected to share $10 evenly with a stranger who had been fair to someone else, when the alternative was to share $12 evenly with an unfair allocator (see our other paper).

A willingness to punish unfairness was also expressed in the telephone surveys. For example, 68 percent of respondents said they would switch their patronage to a drugstore five minutes further away if the one closer to them raised its prices when a competitor was temporarily forced to close; and, in a separate sample, 69 percent indicated they would switch if the more convenient store discriminated against its older workers.

The costs of enforcing fairness are small in these examples—but effective enforcement in the marketplace can often be achieved at little cost to transactors. Retailers will have a substantial incentive to behave fairly if a large number of customers are prepared to drive an extra five minutes to avoid doing business with an unfair firm. The threat of future punishment when competitors enter may also deter a temporary monopolist from fully exploiting short-term profit opportunities.

In traditional economic theory, compliance with contracts depends on enforcement. It is a mild embarrassment to the standard model that experimental studies often produce fair behavior even in the absence of enforcement (Hoffman and Spitzer 1982, 1985; Kahneman, Knatsche, and Thaler 1986; Roth, Malouf, and Murninghan 1981; Reinhard Selten 1978). These observations, however, merely confirm common sense views of human behavior. Survey results indicate a belief that unenforced compliance to the rules of fairness is common. This belief was examined in two contexts: tipping in restaurants and sharp practice in automobile repairs.

Question 17A. If the service is satisfactory, how much of a tip do you think most people leave after ordering a meal costing $10 in a restaurant that they visit frequently?

(N = 122) Mean response = $1.28

Question 17B. In a restaurant on a trip to another city that they do not expect to visit again?

(N = 124) Mean response = $1.27

The respondents evidently do not treat the possibility of enforcement as a significant factor in the control of tipping. Their opinion is consistent with the widely observed adherence to a 15 percent tipping rule even by one-time customers who pay and tip by credit card, and have little reason to fear embarrassing retaliation by an irate server.

The common belief that tipping is controlled by intrinsic motivation can be accommodated with a standard microeconomic model by extending the utility function of individuals to include guilt and self-esteem. A more difficult question is whether firms, which the theory assumes to maximize profits, also fail to exploit some economic opportunities because of unenforced compliance with rules of fairness. The following questions elicited expectations about the behavior of a garage mechanic dealing with a regular customer or with a tourist.

Question 18A. [A man leaves his car with the mechanic at his regular / A tourist leaves his car at a] service station with instructions to replace an expensive part. After the [customer / tourist] leaves, the mechanic examines the car and discovers that it is not necessary to replace the part; it can be repaired cheaply. The mechanic would make much more money by replacing the part than by repairing it. Assuming the [customer / tourist] cannot be reached, what do you think the mechanic would do in this situation?

Make more money by replacing the part
customer: 60% tourist: 63%
Save the customer money by repairing the part
Customer: 40% Tourist: 37%

Question 18B. Of ten mechanics dealing with a [regular customer / tourist], how many would you expect to save the customer money by repairing the part?

Mean response
Customer: 3.62 Tourist: 3.72

The respondents do not approach garages the wide-eyed naive faith. It is therefore all more noteworthy that they expect a tourist and a regular customer to be treated alike, in spite of the obvious difference between the two cases in the potential for any kind of enforcement, including reputation effects.[2]

Here again, there is no evidence that the public considers enforcement a significant factor. The respondents believe that most mechanics (usually excluding their own) would be less than saintly in this situation. However, they also appear to believe that the substantial minority of mechanics who would treat their customers fairly are not motivated in each case by the anticipation of sanctions.

5. Economic Consequences

The findings of this study suggest that many actions that are both profitable in the short run and not obviously dishonest are likely to be perceived as unfair exploitations of market power.[3] Such perceptions can have significant consequences if they find expression in legislation or regulation (Zajac 1978; forthcoming). Further, even in the absence of government intervention, the actions of firms that wish to avoid a reputation for unfairness will depart in significant ways from the standard model of economic behavior. The survey results suggest four propositions about the effects of fairness considerations on the behavior of firms in customer markets, and a parallel set of hypotheses about labor markets.

Fairness in Customer Markets

Proposition 1. When excess demand in a customer market is unaccompanied by increases in suppliers' costs, the market will fail to clear in the short run.

Evidence supporting this proposition was described by Phillip Cagan (1979), who concluded from a review of the behavior of prices that, "Empirical studies have long found that short-run shifts in demand have small and often insignificant effects [on prices]" (p. 18). Other consistent evidence comes from studies of disasters, where prices are often maintained at their reference levels although supplies are short (Douglas Dacy and Howard Kunreuther 1969).

[2] Other respondents were asked to assess the probable behavior of their own garage under similar circumstances: 88 percent expressed a belief that their garage would act fairly toward a regular customer, and 86 percent stated that their garage would treat a tourist and a regular customer similarly.

[3] This conclusion probably holds in social and cultural groups other than the Canadian urban samples studied here, although the detailed rules of fairness for economic transactions may vary.

A particularly well-documented illustration of the behavior predicted in proposition 1 is provided by Alan Olmstead and Paul Rhode (1985). During the spring and summer of 1920 there was a severe gasoline shortage in the U.S. West Coast where Standard Oil of California (SOCal) was the dominant supplier. There were no government-imposed price controls, nor was there any threat of such controls, yet SOCal reacted by imposing allocation and rationing schemes while maintaining prices. Prices were actually higher in the East in the absence of any shortage. Significantly, Olmstead and Rhode note that the eastern firms had to purchase crude at higher prices while SOCal, being vertically integrated, had no such excuse for raising price. They conclude from confidential SOCal documents that SOCal officers "were clearly concerned with their public image and tried to maintain the appearance of being 'fair' " (p. 1053).

Proposition 2. When a single supplier provides a family of goods for which there is differential demand without corresponding variation of inputs costs, shortages of the most valued items will occur.

There is considerable support for this proposition in the pricing of sport and entertainment events, which are characterized by marked variation of demand for goods or services for which costs are about the same (Thaler 1985). The survey responses suggest that charging the market-clearing price for the most popular goods would be judged unfair.

Proposition 2 applies to cases such as those of resort hotels that have in-season and out-of-season rates which correspond to predictable variations of demand. To the extent that constraints of fairness are operating, the price adjustments should be insufficient, with excess demand at the peak. Because naive accounting does not properly distinguish between marginal and average costs, customers and other observers are likely to adopt off-peak prices as a reference in evaluating the fairness of the price charged to peak customers. A revenue-maximizing (low) price in the off-season may suggest that the profits achievable at the peak are unfairly high. In spite of a substantial degree of within-season price variation in resort and ski hotels, it appears to be the rule that most of these establishments face excess demand during the peak weeks. One industry explanation is "If you gouge them at Christmas, they won't be back in March."

Proposition 3. Price changes will be more responsive to variations of costs than to variations of demand, and more responsive to cost increases than to cost decreases.

The high sensitivity of prices to short-run variations of costs is well documented (Cagan 1979). The idea of asymmetric price rigidity has a history of controversy (Kuran 1983; Solow 1980; Stigler and Kindahl 1970) and the issue is still unsettled. Changes of currency values offer a potential test of the hypothesis that cost increases tend to be passed on quickly and completely, whereas cost decreases can be retained at least in part. When the rate exchange between two currencies changes after a prolonged period of stability, the prediction from proposition 3 is that upward adjustments of import prices in one country will occur faster than the downward adjustments expected in the other.

Proposition 4. Price decreases will often take the form of discounts rather than reductions in the list or posted price.

This proposition is strongly supported by the data of Stigler and Kindahl. Casual observation confirms that temporary discounts are much more common than temporary surcharges. Discounts have the important advantage that their subsequent cancellation will elicit less resistance than an increase in posted price. A temporary surcharge is especially aversive because it does not have the prospect of becoming a reference price, and can only be coded as a loss.

Fairness in Labor Markets

A consistent finding of this study is the similarity of the rules of fairness that apply to prices, rents, and wages. The correspondence extends to the economic predictions that may be derived for the behavior of wages in labor markets and of prices in customer markets. The first proposition about prices asserted that resistance to the exploitation of short-term fluctuations of demand could prevent markets from clearing. The corresponding prediction for labor markets is that wages will be relatively insensitive to excess supply.

The existence of wage stickiness is not in doubt, and numerous explanations have been offered for it. An entitlement model of this effect invokes an implicit contract between the worker and the firm. Like other implicit contract theories, such a model predicts that wage changes in a firm will be more sensitive to recent firm profits than to local labor market conditions. However, unlike the implicit contract theories that emphasize risk shifting (Azariadis 1975; Baily 1974; Gordon 1974), explanations in terms of fairness (Akerlof, 1979, 1982; Okun 1981; Solow 1980) lead to predictions of wage stickiness even in occupations that offer no prospects for long-term employment and therefore provide little protection from risk. Okun noted that "Casual empiricism about the casual labor market suggests that the Keynesian wage floor nonetheless operates; the pay of car washers or stock clerks is seldom cut in a recession, even when it is well above any statutory minimum wage" (1981, p. 82), and he concluded that the employment relation is governed by an "invisible handshake," rather than by the invisible hand (p. 89).

The dual-entitlement model differs from a Keynesian model of sticky wages, in which nominal wage changes are always nonnegative. The survey findings suggest that nominal wage cuts by a firm that is losing money or threatened with bankruptcy do not violate community standards of fairness. This modification of the sticky nominal wage dictum is related to proposition 3 for customer markets. Just as they may raise prices to do so, firms may also cut wages to protect a positive reference profit.

Proposition 2 for customer markets asserted that the dispersion of prices for similar goods that cost the same to produce but differ in demand will be insufficient to clear the market. An analogous case in the labor market involves positions that are similar in nominal duties but are occupied by individuals who have different values in the employment market. The prediction is that differences in income will be insufficient to eliminate the excess demand for the individuals considered most valuable, and the excess supply of those considered most dispensable. This prediction applies both within and among occupations.

Robert Frank (1985) found that the individuals in a university who already are the most highly paid in each department are also the most likely targets for raiding. Frank explains the observed behavior in terms of envy and status. An analysis of this phenomenon in terms of fairness is the same as for the seasonal pricing of resort rooms: Just as prices that clear the market at peak demand will be perceived as gouging if the resort can also afford to operate at off-peak rates, a firm that can afford to pay its most valuable employees their market value may appear to grossly underpay their less-valued colleagues. A related prediction is that variations among departments will also be insufficient to clear the market. Although salaries are higher in academic departments that compete with the private sector than in others, the ratio of job openings to applicants is still lower in classics than in accounting.

The present analysis also suggests that firms that frame a portion of their compensation package as bonuses or profit sharing will encounter relatively little resistance to reductions of compensation during slack periods. This is the equivalent of proposition 4. The relevant psychological principle is that losses are more aversive than objectively equivalent foregone gains. The same mechanism, combined with the money illusion, supports another prediction: Adjustments of real wages will be substantially greater in inflationary periods than in periods of stable prices, because the adjustments can then be achieved without making nominal cuts—which are always perceived as losses and are therefore strongly resisted. An unequal distribution of gains is more likely to appear fair than a reallocation in which there are losers.

This discussion has illustrated several ways in which the informal entitlements of customers or employees to the terms of reference transactions could enter an economic analysis. In cases such as the pricing of resort facilities, the concern of customers for fair pricing may permanently prevent the market from clearing. In other situations, the reluctance of firms to impose terms that can be perceived as unfair acts as a friction-like factor. The process of reaching equilibrium can be slowed down if no firm wants to be seen as a leader in moving to exploit changing market conditions. In some instances an initially unfair practice (for example, charging above list price for a popular car model) may spread slowly until it evolves into a new norm—and is no longer unfair. In all these cases, perceptions of transactors' entitlements affect the substantive outcomes of exchanges, altering or preventing the equilibria predicted by an analysis that omits fairness as a factor. In addition, considerations of fairness can affect the form rather than the substance of price or wage setting. Judgments of fairness are susceptible to substantial framing effects, and the present study gives reason to believe that firms have an incentive to frame the terms of exchanges so as to make them appear "fair."

REFERENCES

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, 84: 488–500.

———. 1979. "The Case against Conservative Macroeconomics: An Inaugural Lecture." *Economica*, 46: 219–37.

———. 1980. "A Theory of Social Custom, of Which Unemployment May Be One Consequence." *Quarterly Journal of Economics*, 94: 749–75.

———. 1982. "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics*, 97: 543–69.

Akerlof, George A., and Janet L. Yellen. 1985. "Can Small Deviations from Rationality Make Significant Differences to Economic Equilibrium?" *American Economic Review*, 75: 708–20.

Arrow, Kenneth. 1973. "Social Responsibility and Economic Efficiency." *Public Policy*, 21: 303–17.

Azariadis, Costas. 1975. "Implicit Contracts and Unemployment Equilibria." *Journal of Political Economy*, 83: 1183–202.

Baily, Martin N. 1974. "Wages and Employment Under Uncertain Demand." *Review of Economic Studies*, 41: 37–50.

Bazerman, Max H. 1985. "Norms of Distributive Justice in Interest Arbitration." *Industrial and Labor Relations Review*, 38: 558–70.

Cagan Phillip. *Persistent Inflation: Historical and Policy Essays*. New York: Columbia University Press, 1979.

Dacy, Douglas C., and Howard Kunreuther. 1969. *The Economics of Natural Disasters*. New York: Free Press.

Frank, Robert H. 1985. *Choosing the Right Pond; Human Behavior and the Quest for Status*. New York: Oxford University Press.

Gordon, Donald F. 1974. "A Neo-Classical Theory of Keynesian Unemployment." *Economic Inquiry*, 12: 431–59.

Guth, Werner, Roll Schmittberger, and Bernd Schwarze. 1987. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*, 3: 367–88.

Hoffman, Elizabeth, and Matthew L. Spitzer. 1982. "The Coase Theorem. Some Experimental Tests." *Journal of Law and Economics*, 25: 73–98.

———. 1985. Entitlements, Rights, and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice." *Journal of Legal Studies*, 14: 259–97.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler. 1986. "Fairness and the Assumption of Economics." *Journal of Business*, 59(4): S285–300.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, 47: 263–91.

———. 1984. "Choices, Values, and Frames." *American Psychologist*, 39: 341–50.

Kuran, Timur. 1983. "Asymmetric Price Rigidity and Inflationary Bias." *American Economic Review*, 73: 373–82.

Okun, Arthur. 1981. *Prices and Quantities: A Macroeconomic Analysis*. Washington, D.C.: The Brookings Institution.

Olmstead, Alan L., and Paul Rhode. 1985. "Rationing Without Government: The West Coast Gas Famine of 1920." *American Economic Review*, 75: 1044–55.

Roth, Alvin, Michael Malouf, and J. Keith Murnighan. 1981. "Sociological Versus Strategic Factors in Bargaining." *Journal of Economic Behavior and Organization*, 2: 153–77.

Selten, Reinhard. 1978. "The Equity Principle in Economic Behavior." In *Decision Theory and Social Ethics. Issues in Social Choice*, edited by Hans W. Gottinger and Werner Leinfellner. Dordrecht: D. Reidel.

Solow, Robert M. 1980. "On Theories of Unemployment." *American Economic Review*, 70: 1–11.

Stigler, George J., and James K. Kindahl. 1970. *The Behavior of Industrial Prices*, NBER, New York: Columbia University Press.

Thaler, Richard. 1980. "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization*, 1: 39–60.

———. 1985. "Mental Accounting and Consumer Choice." *Marketing Science*, 4: 199–214.

Tversky, Amos, and Daniel Kahneman. 1986. "Rational Choice and the Framing of Decisions." *Journal of Business*, 59(4): S251–78.

Zajac, Edward E. 1978. *Fairness or Efficiency: An Introduction to Public Utility Pricing*. Cambridge: Ballinger.

———. 1985. "Perceived Economic Justice: The Example of Public Utility Regulation." In *Cost Allocation: Methods, Principles and Applications*, edited by H. Peyton Young. Amsterdam: North-Holland.

C H A P T E R 9

A Theory of Fairness, Competition, and Cooperation

ERNST FEHR AND KLAUS M. SCHMIDT

1. Introduction

Almost all economic models assume that *all* people are *exclusively* pursuing their material self-interest and do not care about "social" goals per se. This may be true for some (may be many) people, but it is certainly not true for everybody. By now we have substantial evidence suggesting that fairness motives affect the behavior of many people. The empirical results of Kahneman, Knetsch, and Thaler (1986), for example, indicate that customers have strong feelings about the fairness of firms' short-run pricing decisions, which may explain why some firms do not fully exploit their monopoly power. There is also much evidence suggesting that firms' wage setting is constrained by workers' views about what constitutes a fair wage (Blinder and Choi 1990; Agell and Lundborg 1995; Bewley 1998; Campbell and Kamlani 1997). According to these studies, a major reason for firms' refusal to cut wages in a recession is the fear that workers will perceive pay cuts as unfair, which in turn is expected to affect work morale adversely. There are also many well-controlled bilateral bargaining experiments which indicate that a non-negligible fraction of the subjects do not care *solely* about material payoffs (Güth and Tietz 1990; Roth 1995; Camerer and Thaler 1995). However, there is also evidence that seems to suggest that fairness considerations are rather unimportant. For example, in competitive experimental markets with complete contracts, in which a well-defined homogeneous good is traded, *almost all* subjects behave as if they are only interested in their material payoff. Even if the competitive equilibrium implies an extremely uneven distribution of the gains from trade, equilibrium is reached within a few periods (Smith and Williams 1990; Roth, Prasnikar, Okuno-Fujiwara, and Zamir 1991; Kachelmeier and Shehata 1992; Güth, Marchand, and Rullière 1997).

There is similarly conflicting evidence with regard to cooperation. Reality provides many examples indicating that people are more cooperative than is assumed

We would like to thank seminar participants at the Universities of Bonn, Berlin, Harvard Princeton and Oxford, the European Summer Symposium on Economic Theory 1997 at Gerzensee (Switzerland) and the ESA conference in Mannheim for helpful comments and suggestions. We are particularly grateful to three excellent referees and to Drew Fudenberg and John Kagel for their insightful comments. The first author also gratefully acknowledges support from the Swiss National Science Foundation (project number 12-43590.95) and the Network on the Evolution of Preferences and Social Norms of the MacArthur Foundation. The second author acknowledges financial support by the German Science Foundation through grant SCHM 119614-1.

in the standard self-interest model. Well-known examples show that many people vote, pay their taxes honestly, participate in unions and protest movements, or work hard in teams even when the pecuniary incentives go in the opposite direction.[1] This is also shown in laboratory experiments (Dawes and Thaler 1988; Ledyard 1995). Under some conditions it has even been shown that subjects achieve nearly full cooperation although the self-interest model predicts complete defection (Isaac and Walker 1988, 1991; Ostrom and Walker 1991; Fehr and Gächter 2000).[2] However, as we will see in more detail in section 4, there are also those conditions under which a vast majority of subjects completely defects as predicted by the self-interest model.

There is thus a bewildering variety of evidence. Some pieces of evidence suggest that many people are driven by fairness considerations, other pieces indicate that virtually all people behave as if completely selfish and still other types of evidence suggest that cooperation motives are crucial. In this chapter we ask whether it is possible to explain this conflicting evidence by a *single simple* model. Our answer to this question is affirmative if one is willing to assume that, in addition to purely self-interested people, there is a *fraction* of people who are also motivated by fairness considerations. No other deviations from the standard economic approach are necessary to account for the evidence. In particular, we do not relax the rationality assumption.[3]

We model fairness as self-centered inequity-aversion. Inequity-aversion means that people resist inequitable outcomes, that is, they are willing to give up some material payoff to move in the direction of more equitable outcomes. Inequity-aversion is self-centered if people do not care per se about inequity that exists among other people but are interested only in the fairness of their own material payoff relative to the payoff of others. We show that in the presence of some inequity-averse people, "fair" and "cooperative" as well as "competitive" and "noncooperative" behavioral patterns can be explained in a coherent framework. A main insight of our examination is that the heterogeneity of preferences interacts in important ways with the economic environment. We show, in particular, that the economic environment determines the preference type that is decisive for the prevailing behavior in equilibrium. This means, for example, that under certain competitive conditions a single purely selfish player can induce a large number of extremely inequity-averse players to behave in a completely selfish manner, too. Likewise, under certain conditions for the provision of a public good, a single selfish player is capable of inducing all other players to contribute

[1] On voting see Mueller (1989). Skinner and Slemroad (1985) argue that the standard self-interest model substantially underpredicts the number of honest taxpayers. Successful team production, for example, in Japanese-managed auto factories in North America, is described in Rehder (1990). Whyte (1955) discusses how workers establish "production norms" under piece-rate systems.

[2] Isaac and Walker and Ostrom and Walker allow for cheap talk, while in Fehr and Gächter subjects could punish each other at some cost.

[3] This differentiates our model from learning models (e.g., Roth and Erev 1995) that relax the rationality assumption but maintain the assumption that all players are interested only in their own material payoff. The issue of learning is further discussed in section 7.

nothing to the public good although the others may care greatly about equity. We also show, however, that there are circumstances in which the existence of a few inequity-averse players creates incentives for a majority of purely selfish types to contribute to the public good. Moreover, the existence of inequity-averse types may also induce selfish types to pay wages above the competitive level. This reveals that, in the presence of heterogeneous preferences, the economic environment has a whole new dimension of effects.[4]

The rest of the paper is organized as follows. In section 2 we present our model of inequity aversion. Section 3 applies this model to bilateral bargaining and market games. In section 4 cooperation games with and without punishments are considered. In section 5 we show that, on the basis of plausible assumptions about preference parameters, the majority of individual choices in ultimatum *and* market *and* cooperation games considered in the previous sections are consistent with the predictions of our model. Section 6 deals with the dictator game and with gift exchange games. In section 7 we compare our model to alternative approaches in the literature. Section 8 concludes the discussion.

2. A Simple Model of Inequity-Aversion

An individual is inequity averse if it dislikes outcomes that are perceived as inequitable. This definition raises, of course, the difficult question how individuals measure or perceive the fairness of outcomes. Fairness judgments are inevitably based on a kind of neutral reference outcome. The reference outcome that is used to evaluate a given situation is itself the product of complicated social comparison processes. In social psychology (Festinger 1954; Stouffer et al. 1949; Homans 1961; Adams 1963) and sociology (Davis 1959; Pollis 1968; Runciman 1966) the relevance of social comparison processes has been emphasized for a long time. One key insight of this literature is that *relative* material payoffs affect people's well-being and behavior. As we will see later, without the assumption that at least for some people relative payoffs matter, it is difficult, if not impossible, to make sense of the empirical regularities observed in many experiments. There is, moreover, direct empirical evidence for the importance of relative payoffs. Agell and Lundborg (1995) and Bewley (1998), for example, show that relative payoff considerations constitute an important constraint for the internal wage structure of firms. In addition, Clark and Oswald (1996) show that comparison incomes have a significant impact on overall job satisfaction. They construct a comparison income level for a random sample of roughly 10,000 British individuals by computing a standard earnings equation. This earnings equation determines the predicted or expected wage of an individual with given socioeconomic characteristics. Then the authors examine the impact of this comparison wage on overall job satisfaction.

[4] Our chapter is, therefore, motivated by a similar concern as the papers by Haltiwanger and Waldman (1985) and Russel and Thaler (1985). While these authors examine the conditions under which nonrational or quasi-rational types affect equilibrium outcomes, we analyze the conditions under which fair types affect the equilibrium.