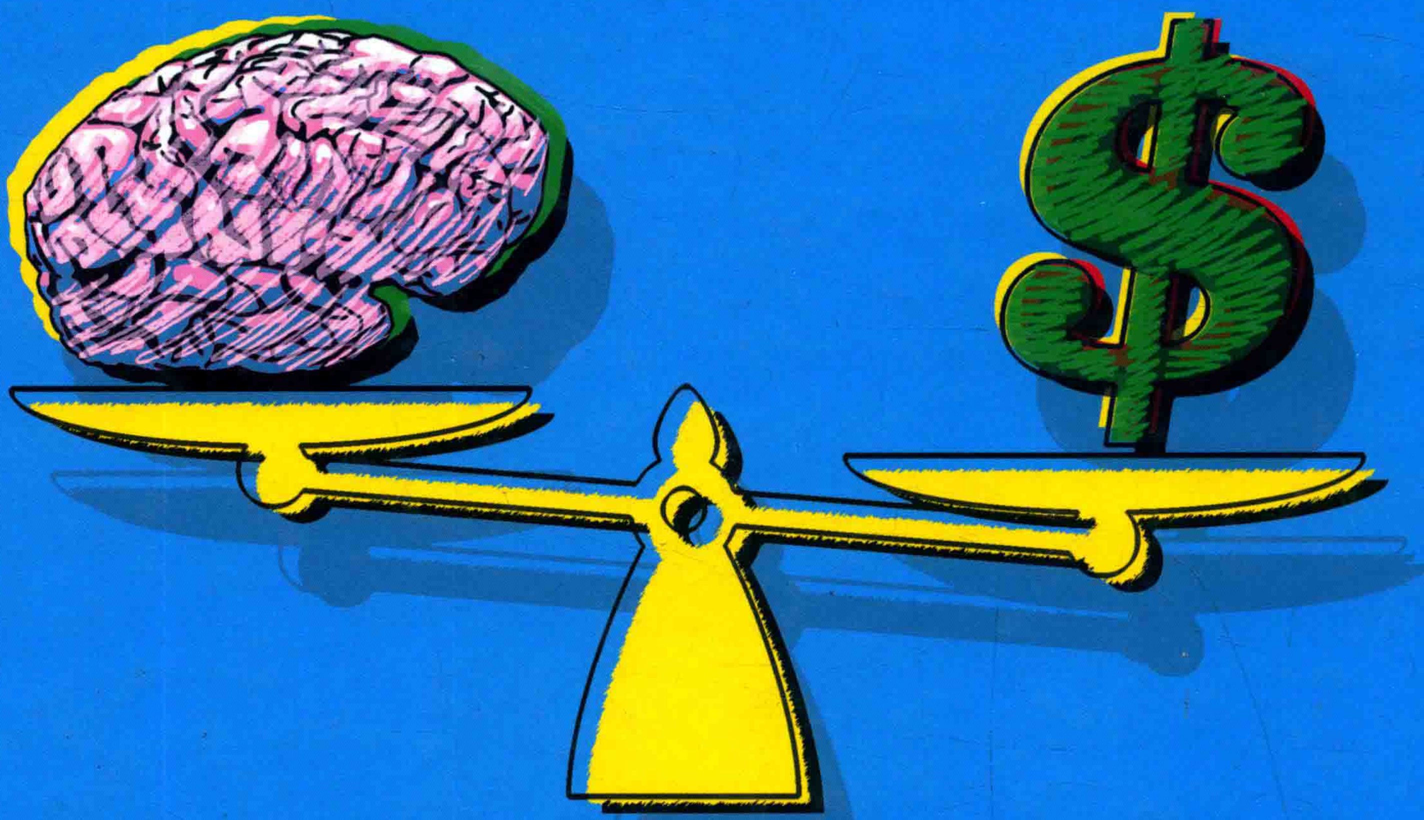
Neuroeconomics

SECOND EDITION

DECISION MAKING AND THE BRAIN



EDITED BY

PAUL W. GLIMCHER • ERNST FEHR

NEUROECONOMICS

Decision Making and the Brain

SECOND EDITION

Edited by

PAUL W. GLIMCHER
New York University, New York, NY, USA

ERNST FEHR
University of Zurich, Switzerland

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

AP

Academic Press is an imprint of Elsevier
32 Jamestown Road, London NW1 7BY, UK
225 Wyman Street, Waltham, MA 02451, USA
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN : 978-0-12-416008-8

For information on all Academic Press publications visit our website at elsevierdirect.com

Typeset by MPS Limited, Chennai, India
www.adi-mps.com

Printed and bound in China

14 15 16 17 10 9 8 7 6 5 4 3 2 1

NEUROECONOMICS

SECOND EDITION

Preface

Since the publication of the first edition of this book, there has been an increasing institutional recognition of the importance of Neuroeconomics in the future of neuroscience, economics, and psychology. At the time that the first edition was published, just a handful of academic institutions included scientists engaged in neuroeconomic research, but the field has matured at an astonishing rate over the past 5 years. Today scholars at nearly a hundred institutions worldwide are at work on neuroeconomic problems and courses on neuroeconomics are now commonplace at both the graduate and undergraduate levels.

At the time that the first edition of this volume was published, it was also true that very little was known about the biological mechanism of human and animal decision making. Accordingly, the first edition was primarily designed to provide scholars interested in beginning to undertake neuroeconomic research with a strong interdisciplinary background in the area. Today, however, the landscape is quite different. Over the past 5 years the field of neuroeconomics has matured intellectually as well as institutionally. Thanks to the work of hundreds of cutting-edge scholars, we now know quite a lot about how and where decisions are made in the brain, and that is reflected in the structure of this second edition.

As we did in the first edition, we continue to believe that a strong interdisciplinary background is important for scholars in this area, but we now confine that review to the first section of the book. The second section presents what is known to date about the neural structure of preferences ranging from risk attitudes to social preferences, and to inter-temporal choice. The third section focuses on the learning of values and the neural systems central to our understanding of the neural representation of subjective value. Section 4 examines what is known about the choice process itself; the mechanism interposed between the valuation processes described in Section 3 and behavior. Section 5 expands on social studies of decision making, an important frontier in neuroeconomic research today. The book concludes with an appendix describing the prospect theory of Kahneman and Tversky in detail. It provides important practical information on the use of prospect theory in neuroeconomic experiments.

USING THE BOOK

One of the critical challenges facing anyone interested in Neuroeconomics is the interdisciplinary nature of the science. To be a neuroeconomist one must be fluent in the languages of Economics, Psychology, Neuroscience, and (to a lesser degree) Primate Anthropology. We recognize that very few scholars or students will come to this book with a background in all of these areas, and the first section of the book is designed to address that constraint before the later sections of the book are encountered.

Section 1

Accordingly, the first section of this book is broken into 4 components. Chapters 1 and 2 describe the theory and methods of experimental economics. For someone trained in economics these chapters would be entirely superfluous. They are intended for neurobiologists and psychologists who are trying to get a handle on economic thought as a preparation for the second through fifth sections of the book. For someone who knows a little economics but who has not been formally trained in economics, these two chapters are essential. Chapters 3 and 4 (along with the appendix) provide a survey of the psychology of Judgment and Decision Making. These chapters should be of particular value to economists and neuroscientists unfamiliar with that tradition. Chapters 5 and 6 are designed to provide basic literacy in the fundamental methodologies of neuroscience for non-neuroscientists. Social scientists who hope to make sense of the empirical chapters that follow are urged to take particular care in reading Chapter 5 which provides a primer on basic neuroscience and Chapter 6 which describes the many methods of neuroscientific research. Even for those non-neuroscientists familiar with basic neuroscience, Chapter 6 should provide a valuable source for understanding the limitations of methods employed in neuroscientific research that can be consulted as one reads the rest of the book. Finally, Chapter 7 provides a useful introduction to the study of non-human animals in decision making. For those whose work, or studies, have focused exclusively on human decision makers, this chapter should provide a clear motivation for understanding the studies of non-human decision makers provided throughout this volume.

For instructors using this volume as a textbook in a graduate or undergraduate class, the first section should typically not be presented in its entirety. If the course is, for example, being presented in a department of psychology as an advanced elective, Chapters 1 and 2 may be all that is required for most students. If the course is for students from many backgrounds, Section 1 may be an appropriate object for self-study early in the class sequence.

Section 2

The second section of the book presents core concepts that guide much neuroeconomic research. The section begins with a first chapter that deals with the basic neural foundations of subjective value in simple binary choice situations and is followed by chapters that engage central notions in neuroeconomic research. Several of the chapters of this section focus on the notion of *preferences;* they describe a core idea in the study of decision making at both a behavioral and a neuroscientific level of analysis. Thus Chapter 9 in this section presents a detailed account of the notion of risk-preference. It describes economic and psychological models of risk preferences that have guided neuroscientific research and then provides a detailed review of current neuroeconomic research in this subarea. Other chapters in this section develop this same theme for intertemporal preferences, social preferences, and for the impact of emotion on preferences. Chapter 13 describes a wealth of research suggesting that neural activity encodes the value of goods and action in a single common neural currency – an idea closely related to the economic notion of utility. Over the last decade, this idea has emerged as a central theme in neuroeconomic research and this chapter reviews those important findings. The section closes with a review of what might be called the *chemistry of choice* (Chapter 14). It examines an emerging thrust of neuroeconomics: the study of how pharmacologic agents like the hormone oxytocin influence choice behavior. To achieve that goal it provides a basic review of neuropharmacology that will be of particular interest to non-neuroscientific readers.

Section 3

The third section of the book focuses on how we learn and represent value. The first chapter in this section (Chapter 15) provides a much-needed overview of the neurobiology of reinforcement learning and dopamine. This is an area where tremendous progress has been made in computational neuroscience and the chapter provides a clear summary of the mathematical and empirical bases for understanding how "values" are learned by the mammalian brain. This will be new material for many economist readers, but even for neurobiologists familiar with studies of the neurotransmitter dopamine, this chapter should provide an important computational foundation. The second chapter of the section, Chapter 16, extends these ideas with a review of advanced topics in reinforcement learning. Together Chapters 15 and 16 should provide an essential starting point for anyone interested in the neural basis of valuation. Chapter 17 builds on these ideas, and requires some familiarity with the material in Chapter 15. It discusses advanced topics in value encoding in other brain areas and presents data on current research frontiers in reinforcement learning. The section concludes with Chapter 18 which presents alternative views of many of the ideas presented in the first three chapters of this section – including important challenges to the core theories presented in the preceding chapters. This chapter will be of particular interest to behavioral economists.

Section 4

The fourth section presents an overview of the choice process: the neural mechanism that takes value or sensory evidence as an input and triggers action as a behavioral output. It begins with two chapters that examine the two main neuroscientific threads in the study of choice: perceptual decision making and value-based decision making. These are followed by a series of chapters that examine advanced topics in this area. Chapter 21 describes evidence that multiple neural systems interact during the choice process. Chapter 22 examines cutting-edge research on the multiple brain systems that integrate costs and benefits in the generation of choice. Chapter 23 describes neuron-level modeling that is beginning to reconcile perceptual, value-based, and cost-related decision making in a single framework. The final chapter in this section examines well-known violations of traditional choice theory and explores the emerging notion that neurobiological constraints may underlie many of these phenomena.

Section 5

This section concludes the volume with an overview of social decision making, within the framework of game theory. Building explicitly on the material presented in Chapter 11 and in Chapter 2, this section of the book explores the neurobiology of social decision making in some detail. It begins with an overview of behavioral game theory in Chapter 25 at the behavioral and neural levels. It then proceeds to studies of the neurobiology of game theoretic behaviors in non-human primates, an area of very active research at this time. It concludes with a review of neurobiological studies of empathy and the theory of mind, which guides neuroeconomic studies of social behavior.

Appendix

The volume concludes with a very detailed appendix on prospect theory. Kahneman and Tversky's prospect theory has played a key role in the development of neuroeconomics, but the theory is more complicated than is generally realized. This chapter thus serves two goals. First, it should provide a detailed and highly valuable how-to guide explaining how prospect theory should, and should not, be used in the laboratory. Second, it provides an overview of current neuroeconomic research on the foundations of prospect theory. Slightly longer than most of the other chapters in the book, it should provide an invaluable hands-on reference for anyone planning prospect theoretic research.

Paul Glimcher and Ernst Fehr

Acknowledgments

We would like to thank the many people whose hard work made this volume possible. First and foremost we want to acknowledge the hard work of the many hundreds of independent researchers whose work is represented in this volume. The authors of the book, a necessarily small subset of those scientists, have worked hard to encapsulate the staggering accomplishments of their peers. They have done what we take to be an amazing job, although we acknowledge that even this huge volume is incomplete because the field of neuroeconomics grows every day. We ask the forgiveness of the many scholars whose work has received inadequate coverage.

We would also like to express particular thanks to the authors of the first edition of this volume. The first edition was much more a survey of neuroeconomics and much less a textbook than is this volume. In order to make it more of a text, we were forced to significantly reduce the number of authors and to sharpen the focus of the book. The authors of the first edition were gracious in allowing us to reuse material from that edition where it seemed appropriate and we here express our gratitude to them for their contribution.

Finally, we wish to express our thanks to those who made this book possible. To Johannes Menzel of Elsevier who was the editor of the first edition, to April Graham who served as the book's associate editor at Academic Press and to Mica Haley at Academic Press who was our editor. Finally, and most importantly, we want to express our truly undying gratitude to one of most important people in the field of Neuroeconomics, Samanta Shaw. Samanta is one of the great heroes of the birth of neuroeconomics, although she is little known outside the core of the field. For the last 5 years she has served as the administrative director of the Society for Neuroeconomics and in that capacity she has probably done more to further the field than anyone else. As the *de facto* editor of this volume she has prepared the second edition (as she did the first edition) through submission, revision, revision again, production, and marketing. We, and all of neuroeconomics, owe her an immeasurable debt. Thanks, Sam.

Paul W. Glimcher and Ernst Fehr

List of Contributors

Adele Diederich Jacobs University, Bremen gGmbH, Bremen, Germany

Andrew Caplin New York University, New York, USA

Anita Tusche Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Antonio Rangel California Institute of Technology, Pasadena, CA, USA; California Institute of Technology, Pasadena, CA, USA

Benedetto De Martino University College London, London, UK; California Institute of Technology Pasadena, USA

Christian C. Ruff University of Zürich, Zürich, Switzerland

Colin F. Camerer California Institute of Technology, Pasadena, CA, USA

Craig R. Fox UCLA Anderson School of Management, Los Angeles, CA, USA

Daeyeol Lee Yale University New Haven, CT, USA

Daniel Houser George Mason University, Fairfax, VA, USA

Elizabeth A. Phelps New York University, New York, NY, USA

Elke U. Weber Columbia University, New York, NY, USA

Eric J. Johnson Columbia University, New York, NY, USA

Ernst Fehr University of Zürich, Zürich, Switzerland

Hauke R. Heekeren Freie Universität Berlin, Berlin, Germany

Hilke Plassmann INSEAD, Fontainebleau, France; Ecole Normale Supérieure, INSERM, Paris, France

Ian Krajbich University of Zürich, Zürich, Switzerland

Jerome R. Busemeyer Indiana University, IN, USA

John A. Clithero California Institute of Technology, Pasadena, CA, USA

John P. O'Doherty California Institute of Technology, Pasadena, CA, USA

Jonathan D. Wallis University of California at Berkeley, Berkeley, CA 94720, USA

Joseph W. Kable University of Pennsylvania, Philadelphia, PA, USA

Joshua I. Gold University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Karolina M. Lempert New York University, New York, NY, USA

Kenji Doya Okinawa Institute of Science and Technology, Okinawa, Japan

Kent C. Berridge University of Michigan, Ann Arbor MI, USA

Kenway Louie New York University, New York, NY, USA

Kevin McCabe Mason University, Arlington, VA, USA

Laurie R. Santos Yale University, New Haven, CT, USA

Matthew F. S. Rushworth University of Oxford, Oxford, UK

Michael C. Dorris Queen's University, Kingston ON, Canada; Chinese Academy of Sciences, Shanghai, China

Michael L. Platt Duke University, Durham, NC, USA; Duke University, Durham, NC, USA

Minoru Kimura Tamagawa University, Machida Tokyo, Japan

Molly J. Crockett University College London, London, UK; University of Zürich, Zürich, Switzerland

Nathaniel D. Daw New York University, New York, USA

Paul W. Glimcher New York University, New York, USA

Phillipe N. Tobler University of Zürich, Zürich, Switzerland

Roger Ratcliff Ohio State University, OH, USA

Russell A. Poldrack University of Texas at Austin, Austin, TX, USA

Scott A. Huettel Duke University, Durham, NC, USA

Tania Singer Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Todd A. Hare University of Zürich, Zürich, Switzerland

Xiao-Jing Wang New York University, New York, NY, USA; NYU Shanghai, Shanghai, China

Yael Niv Princeton University, Princeton, NJ, USA

Introduction: A Brief History of Neuroeconomics

Paul W. Glimcher and Ernst Fehr

With significant material from the first edition by Colin F. Camerer and Russell A. Poldrack

Neuroeconomics has its origins in two places, in events following the neoclassical economic revolution of the 1930s and in the birth of cognitive neuroscience during the 1990s. So we begin this brief history with a review of the neoclassical revolution and the birth of cognitive neuroscience.

NEOCLASSICAL ECONOMICS

The birth of economics is often traced to Adam Smith's publication of *The Wealth of Nations* in 1776. With this publication begins the classical period of economic theory. Smith described a number of phenomena critical to understanding choice behavior and the aggregation of choices into market activity. These were, in essence, psychological insights. They were relatively *ad hoc* rules that explained how features of the environment influenced the behavior of a nation of consumers and producers.

What followed the classical period was an interval during which economic theory became very heterogenous. A number of competing schools with different approaches developed. Many economists of the time (Edgeworth, Ramsey, Fisher) dreamed about tools to infer value from physical signals, through a "hedonimeter" for example, but these early neuroeconomists did not have such tools (Colander, 2007).

One school of thought, due to John Maynard Keynes, was based on the view that regularities in consumer behavior could (among other things) provide a basis for fiscal policy to manage economic fluctuations. Many elements in Keynes' theory such as the "propensity to consume" or entrepreneurs' "animal spirits" that influence their investment decisions were based on psychological concepts. This framework dominated United States fiscal policy until the 1960s.

Beginning in the 1930s a group of economists – most famously, Samuelson, Arrow, and Debreu – began to investigate the mathematical structure of consumer choice and behavior in markets. Rather than simply building models that incorporated a set of parameters that might, on *a priori* psychological grounds, be predictive of choice behavior, this group of theorists began to investigate what mathematical structure of choices might result from simple, more "primitive," assumptions on preferences. Many of these models (and the style of modeling that followed) had a strong *normative*[1] flavor, in the sense that attention was most immediately focused on idealized choices and efficient allocation of resources (as opposed to necessarily seeking to describe how people choose, as psychologists do, and how markets work).

To better understand this approach, consider what is probably the first and most important aspect of these simple models. Instead of assuming that an unknown internal property that we may call "preference" or "utility" causes choices, this approach takes only choices as primitives and subsequently asks whether choices can be *represented* by some form of preference or utility function. In other words, preferences and utility functions have no independent existence from choices. They are just a way of representing or redescribing the object of interest – choices. To see what this means, assume that an individual who can choose between an apple and an orange, chooses the orange. Then we can represent this choice by assigning a higher number to the orange, say 10 for the orange and 9 for the apple and by adding the assumption that higher numbers represent preferred objects. We can also call this a utility function, meaning that U(orange) = 10 and U(apple) = 9. Clearly this utility function represents the individual's choice but so does any other function that assigns a higher number to the orange. It is thus transparent that utility functions only represent choice and have no independent explanatory

[1]In the language of economics one typically distinguishes what are called *positive* and *normative* theories. Positive theories are those which seek only to predict future behavior, either by individuals or by economies. Normative theories are those that seek to make statements about what choices are "best," or in the language of economics *welfare maximizing*. Positive theories are thus *descriptive*, and normative theories *prescriptive*.

power. They are merely a convenient tool for capturing a decision-maker's choices.

The most important consequence of this approach is that it made economics independent from psychology (and neuroscience) because all economists need to study is choices while questions about the neural and psychological processes behind choices are irrelevant as long as human decision making obeys some simple consistency axioms such as the *Weak Axiom of Revealed Preference* (or WARP) or the *Generalized Axiom of Revealed Preferences* (GARP). Both of these axioms are discussed in greater detail in Chapter 1 of this book. GARP essentially means that an individual's choices are transitive while WARP means that if an individual chooses an orange over an apple when both are available in situation 1, then the individual should not choose the apple over the orange in another situation where again both are available. Neoclassical economists such as Samuelson and Houthakker have shown that if individuals obey these consistency axioms powerful implications follow. In particular, their behavior can be represented by the maximization of some utility function, implying that they behave *as if* they maximize some utility function. In addition, the consistency axioms are directly testable at the behavioral level; the existence of utility functions can be subjected to rigorous empirical tests. For example, if individuals violate WARP then their behavior cannot be rationalized by any utility function.

The revealed preference approach, which forms much of the subject matter of Chapters 1 and 2, thus starts from a set of assumptions called axioms which encapsulate a theory of some kind (often a very limited one) in formal language. The poetry in the approach (what distinguishes a beautiful theory from an ugly one) is embodied in the simplicity of the axioms, and the degree to which surprisingly simple axioms make sharp predictions about what kind of choice patterns should and should not be observed. Finally, it is critical to note that what the theory predicts is what new choices could possibly follow from an observed set of previous choices (including choices that respond to policy and other changes in the environment such as responses to changes in prices, taxes, or incomes).

It cannot be emphasized enough how much the revealed-preference view suppressed interest in the psychological nature of preference, because clever axiomatic systems could be used to infer properties of unobservable preference from observable choice (Bruni and Sugden, 2007). Before the neoclassical revolution, Pareto noted in 1897 that:

> It is an empirical fact that the natural sciences have progressed only when they have taken secondary principles as their point of departure, instead of trying to discover the essence of things.... Pure political economy has therefore a great interest in relying as little as possible on the domain of psychology.
>
> Busino (1964; p. xxiv)

The revealed preference revolution in economics achieved Pareto's goal on the basis of clear conceptual foundations and it took several decades until economics returned to psychology as an important source of insights.

What followed the development of WARP and GARP were a series of additional theorems of this type which extended the scope of revealed-preference theory to choices with uncertain outcomes whose probabilities are known (von Neumann and Morgenstern's expected utility theory, EU (1944); Chapter 1) or subjective (or "personal," in Savage's subjective EU theory; SEU; Chapter 9), and in which outcomes may be spread over time (discounted utility theory; Chapter 10). What is most interesting about these theories is that they demonstrate, amongst other things, that a chooser who obeys these axioms must behave both "as if" he has a utility function that constitutes a functional relation between outcomes and their utilities and "as if" his actions were aimed at maximizing total utility. In their seminal book, von Neumann and Morgenstern also laid the foundations for much of *game theory* which they saw as a special problem in utility theory, in which outcomes are generated by the interacting choices of many players.

At the end of this period, neoclassical economics seemed incredibly powerful. Starting with a few simple assumptions which fully described a new theory (for example, expected utility theory), the neoclassicists developed a framework for thinking about and predicting choice. These theories of consumer choice would later form the basis for the demand part of the Arrow–Debreu theory of competitive "general" equilibrium, a system in which prices and quantities of all goods were determined simultaneously by matching supply and demand. This is an important tool because it enables the modeler to anticipate *all* consequences of a policy change: for example, imposing a luxury tax on yachts might increase crime in a shipbuilding town because of a rise in unemployment there. This sort of analysis is unique to economics and partly explains the broad influence of economics in regulation and policymaking.

While the "as if" approach has a thorough conceptual and mathematical foundation, and testable predictions, in the realm of revealed preference theory, Milton Friedman extended the "as if" argument to other domains in a rather questionable way, and without a rigorous foundation. In the 1950s, Friedman wrote an influential book, *The Methodology of Positive Economics*. Friedman argued that assumptions underlying a prediction about market behavior could be wrong, but the prediction could be approximately true. For example, even if a monopolist seller does not sit down with a piece of paper and figure out what price maximizes total profit, monopoly prices might evolve "as if" such a calculation was made (perhaps due to selection pressures within or between firms). Friedman's argument licensed economists to ignore evidence of when economic agents violate rational-choice principles (evidence which is typically from experiments that test the individual choice principles most clearly), a prejudice which is still widespread in economics.

What happened next is critical for understanding where Neuroeconomics came from. The French economist Maurice Allais (1953) designed a series of pairwise choices which led to reliable patterns of revealed preference that violated the central "Independence" axiom of expected utility theory. Allais unveiled his pattern, later called the "Allais paradox," at a conference in France at which many participants, including Savage, the famous founder of subjective expected utility theory, made choices which violated their own theories during an informal lunch. (Savage allegedly blamed the lunchtime wine.)

A few years after Allais's example, Daniel Ellsberg (1961) presented a famous paradox suggesting that the "ambiguity" (Ellsberg's term) or "weight of evidence" (Keynes's term) supporting a judgment of event likelihood could influence choices, violating one of Savage's key axioms. The Allais and Ellsberg paradoxes raised the possibility that the specific properties of EU and subjective EU (SEU) implied by simple preference axioms might be generally wrong. More importantly, the paradoxes invited mathematical exploration (which only began to come to fruition in the 1980s) about how weaker systems of axioms might generalize EU and SEU. The goal of these new theories was to accommodate the paradoxical behavior in a way that is both psychologically plausible and *formally sharp* (by which one means that it does not predict any possible pattern of choices, and could therefore conceivably be falsified by new paradoxes).

One immediate response to this set of observations was to argue that the neoclassical models worked, but only under some limited circumstances, a fact which many of the neoclassicists were happy to concede (for example, Morgenstern once said "the probabilities used must be within certain plausible ranges and not go to .01 or even less to .001"). Surely axioms might also be violated if the details of the options being analyzed were too complicated for the chooser to understand or if the chooser was overwhelmed with too many choices. Observed violations could then be seen as a way to map out boundary conditions – a specification of the kinds of problems that lay outside the limits of the neoclassical framework's range of applicability.

Another hat in the ring was Herbert Simon's suggestion that rationality is computationally bounded, and that much could be learned by understanding "procedural rationality". As a major contributor to cognitive science, Simon clearly had in mind theories of choice which posited particular procedures and suggested the way forward was to understand choice procedures empirically, perhaps in the form of algorithms (of which "always choose the object with the highest utility" is one extreme and computationally demanding procedure.) His notion was, in a sense, that if one understood how the machinery of cognition worked, one would better understand how and why people make the choices they do. And as this book makes clear, formalizing those boundaries is one thread in contemporary neuroeconomics.

A sweeping and constructive alternative view, however, emerged from the work of Daniel Kahneman and Amos Tversky (1979) in the late 1970s and 80s, and other psychologists interested in *judgment and decision making* whose interests intersected with choice theory. What Kahneman, Tversky, and others showed in a series of remarkable experimental examples was that the range of phenomena that fell outside classical expected utility theory was much, much broader than Allais and Ellsberg's examples had suggested. This material forms a principle subject of Chapters 3, 4, and the Appendix.

These psychologists studying the foundations of economic choice found many common choice behaviors – typically easily replicated in experiments – that falsified one or more of the axioms of expected utility theory and which seemed to conflict with fundamental axioms of choice, as described in Chapter 1. For example, some of their experimental demonstrations showed effects of "framing", attacking the implicit axiom of "description-invariance", the idea that choices among objects should not depend on how they are described, a phenomenon described in detail in Chapter 24.

These experiments thus led many scholars, particularly psychologists and economists who had become interested in decision making through the work of Kahneman and Tversky, to conclude that empirical critiques of the simple axiomatic approaches, in the form of counterexamples, could lead to more general axiomatic systems that were more sensibly rooted in principles of psychology.

This group of psychologists and economists, who began to call themselves *behavioral economists,* argued that evidence and ideas from psychology could improve the model of human behavior inherited from neoclassical economics. In one useful definition, behavioral economics proposes models of limits on rational calculation, willpower, and self-interest, and seeks to codify those limits formally and explore their empirical implications using both mathematical theory, experimental data and analysis of field data.

In the realm of risky choice, Kahneman and Tversky modified expected utility to incorporate a psychophysical idea of reference-dependence – valuation of outcomes depend on a point of reference, just as sensations of heat depend on previous temperature – along with a regressive nonlinear transformation of objective probability (details of prospect theory are reviewed in the Appendix, details of the reference point in Chapter 24) Another component of the behavioral program was the idea that statistical intuitions might be guided by *heuristics*, which could be inferred empirically by observing choice under a broad range of circumstances. Heuristics were believed to provide a potential basis for a future theory of choice (see Chapter 3; Gilovich et al., 2002). A third direction are theories of social preference – how people value choices when those choices impact the values of other people (see Chapter 11, and Section 5 of this volume). The goal is to eventually have mathematical systems that embody choice heuristics and specific types of social preference which explain empirical facts but also make sharp predictions. Development of these theories, and tests with both experimental and field data, are now the frontiers of modern behavioral economics.

An obvious conflict developed (and continues to cause healthy debate) between the attempt by behavioral economists to piece together empirically disciplined theories and by the neoclassicists who were arguing for a simpler global theory typically guided by the idea that normative theory is a privileged starting point. The difference in approaches spilled across methodological boundaries too. The influence of ideas from behavioral economics roughly coincided with a rise in interest in conducting carefully controlled experiments on economics systems, by Charles Plott, Vernon Smith, and colleagues (cf. Smith 1976). The experimental economists began with the viewpoint that economic principles should apply everywhere (as principles in natural and physical sciences are presumed to); their view was that when theories fail in simple environments, those failures raise doubt about whether they are likely to work in more complex environments. However, the overlap between behavioral economics and experimental economics is far from complete. Behavioral economics is based on the presumption that incorporating psychological *principles* will improve economic analysis while experimental economics presumes that incorporating psychological *methods* (highly controlled experiments) will improve the testing of economic theory.

In any case, the neoclassical school had clear theory and sharp predictions but the behavioral economists continued to falsify elements of that theory with compelling empirical examples. Neuroeconomics emerged, from within behavioral and experimental economics because behavioral economists often proposed theories that could be thought of as algorithms about both how information was processed, and the choices that resulted from that information processing. A natural step in testing these theories was to simultaneously gather information on both the details of information processing and associated choices. If information processing could be hypothesized in terms of neural activity, then neural measures could be used (along with coarser measures like eyetracking of information to which choosers attend) to test theories as simultaneous restrictions on what information is processed, how that processing works in the brain, and the choices that result. Neuroscientific tools provide further predictions in tests with lesion patient behavior, and transcranial magnetic stimulation (TMS) which should – in theory – change choices if TMS disrupts neural activity in a brain area that is necessary to producing certain kinds of choices. An important backdrop to this development is that economic theorists are extremely clever at inventing multiple systems of axioms which can explain the same patterns of choices. By definition, choices alone provide a limited way to distinguish theories in the face of rapid production of alternative theories. Forcing theories to commit to predictions about underlying neural activity therefore provides a powerful way to adjudicate among theories.

COGNITIVE NEUROSCIENCE

Like economics, the history of the neuroscientific study of behavior also reflects an interaction between two approaches, in this case a neurological approach and a physiological approach. In the standard neurological approach of the last century, human patients or experimental animals with brain lesions were studied in a range of behavioral tasks. The behavioral deficits of the subjects were then correlated with their neurological injuries and the correlation used to infer function. The most classic example of this is probably the work of the British neurologist David Ferrier (1878) who demonstrated that destruction of precentral gyrus of the cortex led to quite precise deficits in movement generation. What marks many of these studies during the classical period in neurology, is that they often focused on damage to either sensory systems or movement control systems. The reason for this should be obvious, the sensory stimuli presented to a subject are easy to control and quantify – they are *observables* in the economic sense of the word. The same is true for movements that we instruct a subject to produce.

Movements are directly observable and easily quantified. In contrast, mental states are much more elusive. Although there has been clear evidence that neurological damage influences mental states for centuries, relating damage to mental state is difficult specifically because mental states are not directly observable. Indeed, relating mental state to neurological damage requires some kind of theory (often a global one), and it was this theory that was largely absent during the classical period in neurology.

In contrast to the neurological approach, the physiological approach to the study of the brain involves correlating direct measurements of biological states, as in the firing of action potentials in neurons, changes in blood flow, and changes in neurotransmitters, with events in the outside world. During the classical period, this more precise set of methodological tools was extremely powerful for elucidating basic features of nervous function but was extremely limited in its applicability to complex mental states. Initially this limitation arose from a methodological constraint. Physiological measurements are invasive, and often destructive. This limits their use to animals and, in the classical period, to anesthetized animals. The result is an almost complete restriction of physiological approaches during the classical period to the study of sensory encoding in the nervous system.

A number of critical advances during the period from the 1960s to the 1980s, however, led both to a broadening of these approaches and later a fusion of these two approaches. Within the domain of neurology, models from psychology began to be used to understand the relationship between brain and behavior. Although the classes of models that were explored were highly heterogeneous, and often not very quantitative, these early steps made it possible to study mental states, at least in a limited way. Within the physiological tradition, technical advances that led to the development of humane methods made it possible to make measurements in awake-behaving animals, also opening the way to the study of mental state, this time in animals.

What followed was a period in which a heterogeneous group of scholars began to develop models of mental processes and then to correlate intermediate variables in these models with either physiological measurements or lesion-induced deficits. But these scholars faced two very significant problems. First, they faced a surplus of models. Dozens of related models could often account for the same phenomena and it was hard to discriminate between models. Second, they faced a paucity of data. Physiological experiments are notoriously difficult and slow and although they yield precise data they do so at an agonizingly slow rate. Neurological experiments (at least in humans) move more quickly but are less precise because the researcher does not have control over the placement of lesions.

It was the resolution of these two problems, or attempts to resolve these problems, which were at the heart of the cognitive neuroscientific revolution. In describing that revolution though, we focus in on the study of decision making. This was by no means a central element in cognitive neuroscientific revolution, but it forms the central piece for understanding the source of neuroeconomics in the neuroscientific community.

The lack of a clear global theory was first engaged seriously by the importation of signal detection theory into the physiological tradition. Signal detection theory (Green and Swets, 1966) is a normative theory of signal categorization broadly used in the study of human perception. The critical innovation that revolutionized the physiological study of cognitive phenomena was the use of this normative theory to relate neuronal activity directly to behavior.

In the late 1980s William Newsome and J. Anthony Movshon (cf. Newsome et al., 1989) began work on an effort to relate the activity of neurons in the middle temporal area of visual cortex (Area MT) to decisions made by monkeys in the domain of perceptual categorization. In those experiments, thirsty monkeys had to evaluate an ambiguous visual signal which indicated which of two actions would yield a fluid reward. What the experiments demonstrated was that the firing rates of single neurons in this area, which were hypothesized to encode the perceptual signal being directly evaluated by the monkeys in their decision making, could be used to predict the patterns of stochastic choice produced by the animals in response to the noisy sensory signals. This was a landmark event in neuroscience because it provided the first demonstration of a clear correlation between neuronal activity and stochastic choice. Following Newsome's suggestion, this class of correlation came to be known as a *psychometric–neurometric match*; the behavioral measurement being referred to as psychometric and the matching neuronal measurement being referred to as neurometric.

This was also a landmark event in the neural study of decision making because it was the first successful attempt to predict decisions from single neuron activity. But it was also controversial. Parallel studies in areas believed to control movement generation (Glimcher and Sparks, 1992) seemed not to be as easily amenable to a signal detection-based analysis (Sparks, 1999; Glimcher, 2003). This led to a long-lasting debate in the early and mid-1990s about whether theories like signal detection would prove adequate for the wholesale study of decision making.

The neurological tradition had gained its first glimpses into the effects of brain damage on decision making in the 1848 case of Phineas Gage (Macmillan, 2002). After his brain was penetrated by a steel rod, Gage exhibited a drastic change in personality and decision making. The systematic study of decision making deficits following brain damage was undertaken beginning in the 1990s by Antonio Damasio, Antoine Bechara and their colleagues (e.g., Bechara et al,, 1994), who began examining decision making under risk in a card sorting experiment. Their work related damage to frontal cortical areas with specific elements of an emotion-based theory of decision making which, though not normative like signal detection theory, was widely influential. The interest in decision making that this work sparked in the neurological community was particularly opportune because at this time the stage was being set for combining a new kind of physiological measurement with behavioral studies in humans.

A better understanding of the relation between mental and neural function in humans awaited the development of methods to image human brain activity non-invasively. Early work by Roland, Raichle, and others had used positron emission tomography (PET) to image the neural correlates to mental function, but this method was limited in its application due to the need for radioactive tracers. In 1992, three groups (Bandettini et al., 1992; Kwong et al., 1992; Ogawa et al., 1992) simultaneously published the first results using functional magnetic resonance imaging (fMRI) to image brain activity non-invasively, a development which opened the door for direct imaging of brain activity while humans engaged in cognitive tasks. This was a critical event because it meant that a technique was available for the rapid (if crude) measurement of neural state in humans directly. Because of the wide availability of MRI and the safety of the method, the use of fMRI for functional imaging of human cognitive processes has grown exponentially. Perhaps because of the visually compelling nature of the results, showing brain areas "lighting up," this work became highly influential not just in the neuroscientific and psychological communities, but also beyond. The result was that scholars in many disciplines began to consider the possibility of measuring the brain activity of humans during decision making. The challenge was that there was no clear theoretical tool for organizing this huge amount of information.

SETTING THE STAGE FOR NEUROECONOMICS

By the late 1990s several converging trends set the stage for the birth of neuroeconomics. Within economics and the psychology of judgment and decision making, a critical tension had emerged between the neoclassical/revealed preference school and the behavioral school. The revealed preference theorists had an elegant axiomatic model of human choice which had been revealed to be only crudely predictive of human behavior and for which it is easy to produce counter-examples. Revealed preference theorists responded to this challenge by both tinkering with the model to improve it and challenging the significance of many of the existing behavioral economic experiments.

The behavioral economists, in contrast, responded to this challenge by looking for alternative mathematical theories and different types of data to test those theories, theories which they saw as claims about both computational/psychological processes and choices. Their goal was to provide an alternative theoretical approach for predicting behavior and a methodology for testing those theories. This is an approach that benefits from good theories that predict both choices and "non-choice" data. The appropriate form for such an alternative theory has been, however, hotly debated. One approach to developing such a theory derives from the great progress economics has made towards understanding the interaction of two agent systems in the external world, for example understanding the interactions of firms and the workers they hire. This preexisting mathematical facility with two-agent models aligned naturally with an interest among psychologists in what are known as "dual process" models. If, as some behavioral economists have argued, the goal is to minimally complicate the standard models from economics, going from viewing a person as a single agent maximizing a unifying "utility" to viewing her as two independent agents (or to two processes) interacting, might be a useful strategy, and this strategy forms one of the principle alternative theoretical approaches that gave birth to neuroeconomics. The appeal of the dual-process model for economists is that when inefficient choice behaviors are observed in humans, these can be viewed as the result of the two (or more) independent agents locked in a bad equilibrium by their own self-interests. Of course, other scholars within behavioral economics have suggested other approaches that also have neuroeconomic implications. A view from evolutionary psychology that may serve as another example is that encapsulated models execute heuristics that are specially adapted to evolutionarily selected tasks (e.g., Gigerenzer *et al.*, 2000). These models have something to say about the tradeoff between efficient choice and computational complexity, which might be used to generate hypotheses about brain processes (and cross-species comparisons).

Within much of neuroscience, and that fraction of cognitive psychology closely allied with animal studies of choice, a different tension was being felt at the same time that these multiple agent and heuristic models were evolving in behavioral economics. It was clear that both those physiologists interested in single neuron studies of decision making, and those cognitive neuroscientists closely allied to them, were interested in describing the algorithmic mechanisms of choice. Their goal was to describe the neurobiological hardware that supported choice behavior in situations ranging from perceptual decision making to the expression of more complicated preferences. What they lacked was an overarching theoretical framework for placing their neural measurements into context. Newsome and his colleagues had argued that the standard mathematical tool for understanding sensory categorization, signal detection theory, could serve that role but many remained skeptical that this approach could be sufficiently generalized. What that naturally led to was the suggestion, by Glimcher and his colleagues, that the neoclassical/revealed preference framework might prove a useful theoretical tool for neuroscience. What followed was the rapid introduction of concepts like expected value and expected utility to the neuroscientific literature.

TWO TRENDS, ONE GOAL

The birth of neuroeconomics, then, grew from a number of related factors that simultaneously influenced what were basically two separate communities, although communities with significant overlap. A group of behavioral economists and cognitive psychologists looked towards functional brain imaging as a tool to both test and develop alternatives to neoclassical/revealed preference theories (especially when too many theories chase too few data using choices as the only class of data). A group of physiologists and cognitive neuroscientists looked towards economic theory as a tool to test and develop algorithmic models of the neural hardware for choice. The result was an interesting split that persists in neuroeconomics today – and of which there is evidence in this volume.

The result is that the two communities, one predominantly (although certainly not exclusively) neuroscientific and the other predominantly (although not exclusively) behavioral economic, thus came towards a union from two very different directions. Both, however, promoted an approach that was controversial within their parent disciplines. Many neurobiologists outside the emerging neuroeconomic community argued that the complex normative models of economics would be of little value for understanding the behavior of real humans and animals. Many economists, particularly hardcore neoclassicists, argued that algorithmic-level studies of decision making were unlikely to improve the predictive power of the revealed preference approach.

Despite these challenges, the actual growth of neuroeconomics during the late 1990s and early 2000s was explosive – and it led to the much more heterogeneous mix of scholars who call themselves neuroeconomists today. The converging group of like-minded economists, neuroscientists, and cognitive psychologists quickly generated a set of meetings and conferences that fostered a growing sense of interdisciplinary collaboration. Probably the first of these interdisciplinary interactions was held in 1997 at Carnegie–Mellon University, organized by the economists Colin Camerer and George Loewenstein. After a hiatus of several years this was followed by two meetings in 2001, one held by the Gruter Foundation for Law at their annual meeting in Squaw Valley. At that meeting the Gruter Foundation chose to focus its workshop on the intersection of neuroscience and economics and invited several speakers active at the interface of these converging disciplines. The second meeting was focused more directly on what would later become Neuroeconomics and was held at Princeton University. The meeting was organized by the neuroscientist Jonathan Cohen and the economist Christina Paxson. This meeting is often seen as the inception point for the present-day Society for Neuroeconomics. At that meeting, economists and neuroscientists met to explicitly discuss the growing convergence of these fields and to explicitly debate the value of such a convergence. There was, however, no consensus that the growing convergence was desirable at that meeting.

Nonetheless, the Princeton meeting generated significant momentum and the following year (2003) a small invitation-only meeting that included nearly all of the active researchers in the emerging area was held on Martha's Vineyard, organized by Greg Berns of Emory University. This three-day meeting marked a clear turning point at which a group economists, psychologists and neurobiologists began to identify themselves as Neuroeconomists and began to explicitly shape the convergence between the fields. This led to an open registration meeting the following year at Kiawah Island organized by Read Montague, then at Baylor College of Medicine. A decision was made at that meeting by essentially all of the central figures in the emerging discipline to form a society and to turn this recurring meeting into an annual event that would serve as a focal point for neuroeconomics internationally. At that meeting Paul Glimcher was elected President of the Society. The Society then held its first formal meeting the following year (2005) at Kiawah Island.

Against this backdrop of meetings, a series of critical papers and books were emerging that did even more to shape these interactions between scholars in the several disciplines and to communicate the goals of the emerging Neuroeconomic community to the larger neurobiological and economic communities. Probably the first neurobiological paper to explicitly rest on a normative economic theory was Peter Shizgal and Kent Conover's 1996 review in *Current Directions in Psychological Science*: "On the Neural Computation of Utility" in which they attempted to describe the neurobiological substrate of a behavioral choice using a form of normative choice theory derived from economics. What Shizgal's work did not do, however, was to fully incorporate the standard economic model.

In 1999 this was followed by a paper by Platt and Glimcher in *Nature* that argued quite explicitly for a normative utility-based analysis of choice behavior in monkeys. As they put it in that paper:

> Neurobiologists have begun to focus increasingly on the study of sensory-motor processing, but many of the models used to describe these processes remain rooted in the classic reflex…Here we describe a formal economic-mathematical approach for the physiological study of the sensory-motor process, or decision-making.

At an experimental level, the paper goes on to demonstrate that the activity of single neurons in the posterior parietal cortex is a lawful function of both the probability and magnitude of expected rewards. This was significant because standard expected utility theory predicates choice on lawful functions of these same two variables.

At the same time that this paper appeared in print the behavioral economists Colin Camerer, George Loewenstein, and Drazen Prelec began circulating a manuscript in economic circles by the name of *Grey Matters*. In this manuscript the authors also argued for a neuroeconomic approach, but this time from a behavioral economic perspective. What these three economists argued was that the failures of traditional axiomatic approaches likely reflected neurobiological constraints on the algorithmic processes responsible for decision making. Neurobiological approaches to the study of decision, they argued, might reveal and define these constraints which cause deviations in behavior from normative theory.

What was striking about this argument, in economic circles, was that it proposed an algorithmic analysis of the physical mechanism of choice, a possibility that has been explicitly taboo until that time. Prior to the 1990s it had been a completely ubiquitous view in economic circles that models of behavior, like expected utility theory, were *as if* models; the model was to be interpreted "as if" utility were represented internally by the chooser. But as Samuelson had argued a half century earlier, it was irrelevant whether this was actually the case because the models sought to link options to choices *not* to make assertions about the mechanisms by which that process was accomplished. Camerer and colleagues argued against this view, suggesting that deviations from normative theory should be embraced as clues to the underlying neurobiological basis of choice. In a real sense then, these economists turned to neurobiology for exactly the opposite reason that the neurobiologists had turned to economics. They embraced neuroscience as a principled alternative to normative theory.

At this point, there was a rush by several research groups to perform an explicitly economic experiment that would mate these two disciplines in human choosers. Two groups succeeded in this quest in 2001. The first of these papers appeared in the journal, *Neuron*, and reflected a collaboration between the functional magnetic resonance imaging pioneer Hans Breiter, Shizgal, and Kahneman (who would win the Nobel prize for his contribution to behavioral economics the following year; Breiter et al., 2001). That paper was based on Kahneman and Tversky's *prospect theory*, a non-normative form of expected utility theory, described in Chapter 3 and the Appendix, that guided much research in judgement and decision-making laboratories throughout the world. In that paper, Breiter and colleagues manipulated the perceived desirability of a particular lottery outcome (in this case winning zero dollars) by changing the values of two other possible lottery outcomes. When winning zero dollars is the worst of three possible outcomes, Kahneman and Tversky's prospect theory predicts that the subjects should view it negatively, but when it is the best of the three outcomes then subjects should view it more positively. The scanning experiment revealed that brain activation in the ventral striatum matched these predicted subjective valuations.

The other paper published that year reflected a collaboration between the more neoclassically oriented economist Kevin McCabe, his colleague Vernon Smith (who would share the Nobel prize with Kahneman the following year for his contributions to experimental economics), the econometrician Daniel Houser and a team that included a psychologist and a biomedical engineer. Their paper, which appeared in the *Proceedings of the National Academy of Sciences* in the United States (McCabe *et al.*, 2001) examined behavior and neural activation while subjects engaged in a strategic game. This also represented the first use of game theory, an economic tool for the study of social decision making, in a neurobiological experiment. In that paper, subjects played a trust game either against