CHAPTER

1

Introduction: A Brief History of Neuroeconomics

Paul W. Glimcher, Colin F. Camerer, Ernst Fehr, and Russell A. Poldrack

OUTLINE

Neoclassical Economics	1	Two Trends, One Goal	7
Cognitive Neuroscience	5	Summary	11
Setting the Stage for Neuroeconomics	6	References	11

Over the first decade of its existence, neuroeconomics has engendered raucous debates of two kinds. First, scholars within each of its parent disciplines have argued over whether this synthetic field offers benefits to their particular parent discipline. Second, scholars within the emerging field itself have argued over what form neuroeconomics should take. To understand these debates, however, a reader must understand both the intellectual sources of neuroeconomics and the backgrounds and methods of practicing neuroeconomists.

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. We therefore 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 began the classical period of economic theory. Smith described a number of phenomena critical for 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, 2008).

One school of thought, due to John Maynard Keynes, was 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 (see, for example, Samuelson, 1938). 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 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 of these simple models: the *Weak Axiom of Revealed Preference* (WARP). WARP was developed in the 1930s by Paul Samuelson, who founded the revealed preference approach that was the heart of the neoclassical revolution. Samuelson proposed that if a consumer making a choice between an apple and an orange selects an apple, he *reveals a preference* for apples. If we assume only that this means he *prefers* (preference is here a stable internal property that economists did not hope to measure directly) apples to oranges, what can we say about his future behavior? Can we say anything at all?

What Samuelson and later authors showed mathematically was that even simple assumptions about binary choices, revealing stable (weak) preferences, could have powerful implications. An extension of the WARP axiom called GARP (the "generalized" axiom of revealed preference, Houthakker, 1950) posits that if apples are revealed preferred to oranges, and oranges are revealed preferred to peaches, then apples are "indirectly" revealed preferred to peaches (and similarly for longer chains of indirect revelation). If GARP holds for binary choices among pairs of objects, then some choices can be used to make predictions about the relative desirability of pairs of objects that have never been directly compared by the consumer. Consider a situation in which a consumer chooses an *apple* over an *orange* and then an *orange* over a *peach*. If the assumption of GARP is correct, then this consumer must not choose a *peach* over an *apple* even if this is a behavior we have never observed before.

The revealed preference approach 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 theory tells us what a series of observed choices implies about intermediate variables such as utilities (and, in more developed versions of the theory, subjective beliefs about random events). 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 which 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). The theories do not predict intermediate variables; they use them as tools. What revealed preference theories predict is choice. It is the only goal, the only reason for being, for these theories.

What followed the development of WARP were a series of additional theorems of this type which extended the scope of revealed-preference theory to choices with uncertain outcomes whose likelihoods are known (von Neumann and Morgenstern's expected utility theory, EU) or subjective (or "personal," in Savage's subjective EU theory), and in which outcomes may be spread over time (discounted utility theory) (see Chapter 3 for more details). 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 continuous utility function that relates the subjective value of any gain to its objective value and "as if" his actions were aimed at maximizing total obtained 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 choices of many players (von Neumann and Morgenstern, 1944).

At the end of this period, neoclassical economics seemed incredibly powerful. Starting with as few as one and as many as four simple assumptions which fully described a new 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 policy-making.

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.
>
> (Quoted in Busino, 1964: xxiv)

Later, in the 1950s, Milton 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 has been 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 that typically comes from experiments that test the individual choice principles most clearly), a prejudice that is still widespread in economics.

What happened next is critical for understanding where neuroeconomics arose. In 1953, the French economist Maurice Allais 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, made choices which violated their own theories during an informal lunch. (Savage allegedly blamed the lunchtime wine.)

A few years after Allais' example, Daniel Ellsberg (1961) presented a famous paradox suggesting that the "ambiguity" (Ellsberg's term) or "weight of evidence" (Keynes' 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 functional forms of EU and subjective EU implied by simple axioms of preference were generally wrong. More importantly, the paradoxes invited mathematical exploration (which only came 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 (i.e., which does not predict that any pattern of choices is possible, 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 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 approach 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 that 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).

A sweeping and constructive view emerged from the work of Daniel Kahneman and Amos Tversky (1979) in the late 1970s and 1980s, 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 even broader than Allais' and Ellsberg's examples had suggested.

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. 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.

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 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 depends on a point of reference, just as sensations of heat depend on previous temperature – along with a regressive non-linear transformation of objective probability. (Details of prospect theory are reviewed in Chapter 11.) 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 (Gilovich *et al.*, 2002). A third direction is theories of social preference – how people value choices when those choices impact the values of other people (see Chapter 15). The goal is eventually to 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 behavioral economists, who were attempting to piece together empirically disciplined theories, and 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 among economists such as Charles Plott, Vernon Smith and colleagues in conducting carefully controlled experiments on economics systems (see, for example, 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 a 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 regarding how information was processed, and the choices that resulted from that information-processing. A natural step in testing these theories was simultaneously to gather information on the details of both 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 that choosers attend to) 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 an 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 classic example of this is probably the work of the British neurologist David Ferrier (1878), who demonstrated that destruction of the 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 state is much more elusive. Although there has for centuries been clear evidence that neurological damage influences mental state, relating damage to mental state is difficult specifically because mental state is 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 state, such as 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 in animals and, in the classical period, in anesthetized animals. The result was 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 to both 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 state, 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 correlate intermediate variables in these models with either physiological measurements or lesion-induced deficits. However, these scholars faced two very significant problems. First, there was a surplus of models. Dozens of related models could often account for the same phenomena, and it was hard to discriminate between these models. Second, there was 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 them, that was at the heart of the cognitive neuroscientific revolution. In describing that revolution, we focus on the study of decision making. This was by no means a central element in the 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 (see, for example, 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 really clear demonstration of a 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 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. However, 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 regarding whether theories such as 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 1848, in the 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 ability. The systematic study of decision-making deficits following brain damage was initially undertaken, in the 1990s, by Antonio Damasio, Antoine Bechara, and their colleagues (see, for example, 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 owing 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 that 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) direct measurement of neural state in humans. Owing to 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 possibilities 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 had 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 was easy to produce counterexamples. 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 (relying on the Friedman "F-twist" – that predictions based on axioms might be approximately true even if the axioms are wrong).

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 being claims about both computational 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 requires good theories that predict both choices and "non-choice" data. The appropriate form for such an alternative theory has, however, been 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 pre-existing 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, then going from a single agent maximizing a unifying "utility" to two independent agents (or processes) interacting might be a useful strategy. 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 being 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 (see, for example, 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 simultaneously being felt as 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 to the neuroscientific literature of such concepts as expected value and expected utility.

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, albeit with a 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 chased 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 not exclusively) neuroscientific and the other predominantly (although not exclusively) behavioral economic, thus approached 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. 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 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, and is often seen as having been the inception of the present-day Society for Neuroeconomics. At this meeting, economists and neuroscientists met to explicitly discuss the growing convergence of these fields and to debate the value of such a convergence. There was, however, no consensus at the meeting that the growing convergence was desirable.

Nonetheless, the Princeton meeting generated significant momentum, and in 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 of economists, psychologists, and neurobiologists began to identify themselves as neuroeconomists and to explicitly shape the convergence between the fields. This led to an open registration meeting the following year at Kiawah Island, organized by Baylor College of Medicine's Read Montague. At this meeting a decision was made, by essentially all 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 the meeting, Paul Glimcher was elected President of the Society. The Society then held its first formal meeting in 2005 at Kiawah Island.

Against this backdrop of meetings, a series of critical papers and books was 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 rest explicitly on a normative economic theory was Peter Shizgal and Kent Conover's 1996 review, "On the neural computation of utility," in *Current Directions in Psychological Science*. This was followed the next year by a related paper published by Shizgal in *Current Opinion in Neurobiology* entitled "Neural basis of utility estimation." The reason that these papers can be viewed as the first in neuroeconomics is because they attempt to describe the neurobiological substrate of a behavioral choice using a form of normative choice theory derived from economics. In these papers, Shizgal analyzed the results of studies of intracranial self-stimulation in rats using a type of utility theory related loosely to the standard expected utility theory of von Neumann and Morgenstern. The papers argue that the choices an animal makes regarding whether or not to work for electrical stimulation of the medial forebrain bundle can be construed as an effort to maximize the animal's instant-to-instant utility. In this analysis, then, changes in the desirability of brain-stimulation reward as a function of stimulation frequency should be formally interpreted as changes in the utility of stimulus train. Unlike in standard theories of utility, however, Shizgal and Conover proposed that the expected utility of an action is perceived by the animal as the expected utility of that action divided by the sum of the expected utilities of all available actions. This particular formulation has its root in the work of the psychologist Richard Herrnstein, who proposed that many choices reflect this normalization with regard to the value of other alternatives – a phenomenon he referred to as *the matching law*. (For more about the matching law, see Chapter 30).

In fact, this equation had been introduced to self-stimulation studies five years earlier by Shizgal's mentor, C. Randy Gallistel. In the early 1990s, Gallistel had used Herrnstein's work to inspire quantitative choice-based experiments and analyses of intracranial self-stimulation (see Gallistel, 1994). Shizgal's extension of this work is critical in the history of neuroeconomics, because he moved away from the largely descriptive models of Herrnstein towards the normative models of economics. What Shizgal's work did not do, however, was fully incorporate the standard economic model, but rather a more normative version of Herrnstein's approach.

In 1999 this set of papers was followed by a paper by Platt and Glimcher (another student of Gallistel's) in *Nature* that argued quite explicitly for a normative utility-based analysis of choice behavior in monkeys (Platt and Glimcher, 1999). 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 the magnitude of expected rewards. This was significant, because standard expected utility theory predicates choice on lawful functions of these same two variables. The paper, however, makes a critical mis-step in its examination of actual choice behavior. The authors go on to examine a matching-law type behavior which they interpret in terms of normative expected utility theory. This is problematic, because there is no normative standard for the analysis of matching-law behaviors. Indeed, in the example

they present in the paper it cannot be proved that the behavior is predicted by their normative model; if anything, the data seem to suggest that the animals' behave sub-optimally. The result is a mixing of normative and non-normative approaches that characterized the early neurobiological work with economic approaches.

At the same time that this paper appeared in print, the behavioral economists Colin Camerer, George Lowenstein, 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 had 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. However, as Samuelson had argued half a 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 in Economic Sciences for his contribution to behavioral economics the following year). This paper (Breiter *et al.*, 2001) was based on Kahneman and Tversky's *prospect theory*, a non-normative form of expected utility theory that guided much research in judgment and decision-making laboratories throughout the world (a theory described in detail in Chapter 11). In the 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 subjects should view it negatively; however, 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 of the United States of America* (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 this paper, subjects played a trust game either against an anonymous human opponent or against a computer, the details of which are reviewed in Chapter 5 of this volume. Their neurobiological data revealed that in some subjects the medial prefrontal cortex is differentially active under some of the conditions they examined, becoming more active when subjects play a cooperative strategy that deviates from the standard normative prediction of play in that game. From these data, the authors hypothesized that this non-normative pattern of cooperation has its origin in circuits of the prefrontal cortex.

The following year, many of these emerging trends were reviewed in an important special Society for Neuroscience conference issue of the journal *Neuron* (Volume 36, Issue 2) edited by Jonathan Cohen and Kenneth Blum entitled *Reward and Decision*. As these editors wrote in the introduction to that issue:

> Within neuroscience, for example, we are awash with data that in many cases lack a coherent theoretical understanding (a quick trip to the poster floor of the Society for Neurosciences meeting can be convincing on this point). Conversely, in economics, it has become abundantly evident that the pristine assumptions of the "standard economic model" – that individuals operate as optimal decision makers in maximizing utility – are in direct violation of even the most basic facts about human behavior.

In that issue, although all of the articles are by neurobiologists, particular attention is drawn to normative theories of decision. Of especial interest are articles by Montague and Berns (2002), Schultz (2002), Dayan and Balleine (2002), Gold and Shadlen (2002), and Glimcher (2002), which all point towards the interaction of normative models and neurobiology. Interestingly, the issue draws attention to the ongoing debate regarding the role of the neurotransmitter dopamine in reward processing, and draws upon previous work that had identified normative or near-normative models of learning that posit a role for dopamine. (This is a subject of tremendous importance to neuroeconomists today, and forms the focus of the third section of this volume.) What followed was a literal flood of decision-making studies in the neuroscientific literature, many of which relied on normative economic theory. Figure 1.1 documents this flood, plotting the number of papers published from 1990 to 2006 that list both "brain" and "decision making" as keywords.

At the end of this initial period, a set of summary reviews began to emerge that served as manifestos for the emerging neuroeconomic discipline. In 2003 Glimcher published a book, directed primarily at neuroscientists, that reviewed the history of neuroscience and argued that this history was striking in its lack of normative models for higher cognitive function (Glimcher, 2003). Glimcher proposed that economics could serve as the source for this much needed normative theory. Shortly thereafter the Camerer, Loewenstein, and Prelec paper was published under the title "Neuroeconomics" (Camerer *et al.*, 2005); this also served as a manifesto, but from the economic side.

Within the economic community a role similar to that of the *Neuron* special issue was played by a special issue on neuroeconomics presented by the journal *Games and Economic Behavior* (Volume 52, Issue 2) and edited by the economist Aldo Rustichini, which appeared shortly after this in 2005. Within the economic community this issue was hugely influential and served, to a large degree, to define neuroeconomics. The issue included articles by several economists and neuroscientists, including scholars ranging from Gallistel (2005) to Smith (Houser *et al.*, 2005).

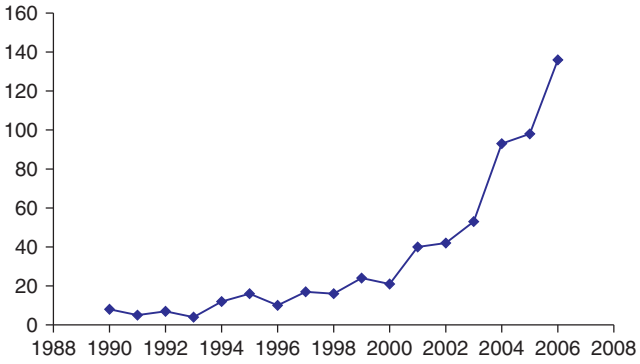

FIGURE 1.1 The increase in numbers of papers on decision-making studies in the neuroscientific literature, 1990–2006

Another major advance was presented in 2005, this one by Michael Kosfeld and his colleagues in Ernst Fehr's research group at the University of Zurich (Kosfeld *et al.*, 2005). This paper was important because it was the first demonstration of a neuropharmacological manipulation that alters behavior in a manner that can be interpreted with regard to normative theory. In the paper, subjects were asked to play a trust game much like the one examined by McCabe and colleagues. Fehr's critical manipulation was to increase brain levels of the neuropeptide oxytocin (by an intranasal application of the compound) before the players made their decision. What Kosfeld and colleagues found was that the investors with oxytocin sent more money to the trustees in the trust game than investors who received placebo. This increase in trusting behavior occurred despite the fact that investors' beliefs about the trustees' back-transfers remained unchanged. In contrast, oxytocin did not affect the trustees' behavior – i.e., trustees' back-transfers remained unchanged – ruling out the possibility that the neuropeptide just increases reciprocity or generosity. However, oxytocin did not cause an unspecific increase in the willingness to take risks, because in a control experiment – a pure risk game – the investors with oxytocin did not behave differently from the subjects with placebo. What was most interesting about this study from a neuroeconomic point of view was the demonstration that the administration of this endogenously produced neuropeptide altered a complex choice behavior of subjects in a very specific way – it neither affected the trustees' behavior nor did it affect the investors' general willingness to take risks, it only increased the investors' risk preference if the risk was constituted by the interaction with another human partner – suggesting a neurobiological basis for a difference between preferences for social and non-social risks.

The rise of neuroeconomics has been strongly associated with the rapid development of non-invasive neuroimaging techniques for human research and single-cell recordings in non-human primates. One limitation of these technologies is that they produce largely correlative measures of brain activity, making it difficult to examine the causal role of specific brain activations for choice behavior. This limitation can, however, be overcome with non-invasive methods of brain stimulation, such as transcranial magnetic stimulation (TMS) and transcranial direct current

stimulation (tDCS), which enable researchers selectively to modify the neural processing associated with choice behavior. A recent study by Knoch *et al.* (2006) provides a demonstration of the additional neuroeconomic insights generated with these methods. Previous fMRI results (Sanfey *et al.*, 2003) had shown that the right *and* the left dorsolateral prefrontal cortex (DLPFC) are activated when subjects decide about the acceptance or rejection of unfair bargaining offers in the ultimatum game (for a description of this bargaining game, see Chapter 5). This finding raises many points, such as whether both hemispheres are causally involved in the choice process. Likewise, is DLPFC affecting judgments about the fairness of bargaining offers, or is it specifically involved in the implementation of fairness concerns? Knoch and colleagues disrupted the right and the left DLPFC with TMS and found that the disruption of both PFC areas left more abstract judgements of fairness fully intact (relative to a placebo stimulation), while the disruption of the right (but not the left) DLPFC resulted in a large increase in the acceptance of unfair offers. From a neuroeconomic viewpoint it is important to know the dissociations between judgment and choice, because choice typically implies that the decision maker must bear costs and benefits, while judgment alone is not yet associated with the bearing of costs and benefits. More generally, non-invasive brain stimulation techniques are likely to play an important role in future neuroeconomic studies because they provide causal knowledge and, in combination with imaging tools, make it possible to isolate whole decision networks that are causally involved in the generation of choices.

SUMMARY

Despite these impressive accomplishments, neuroeconomics is at best a decade old and has yet to demonstrate a critical role in neuroscience, psychology, or economics. Indeed, scholars within neuroeconomics are still debating whether neuroscientific data will provide theory for economists or whether economic theory will provide structure for neuroscience. We hope that both goals will be accomplished, but the exact form of this contribution is not yet clear. However, there are also skeptical voices, and the Pareto (1897) and Friedman arguments that economics is only about choices still lives in the form of fundamentalist critique. Gul and Pesendorfer (2008), for example, have argued that neuroscientific data and neuroscientific theories should, in principle, be unwelcome in economics.

The chapters that follow should allow readers to draw their own conclusions regarding this growing and dynamic field. Each of the major threads of contemporary research is reviewed in these pages. Although it is far too soon for there to be consensus in this community, the field today is small enough that a single volume can provide a comprehensive review. We therefore invite you, the readers, to estimate for yourselves the future directions that will yield greatest profit.

References

Allais, M. (1953). Le comportment de l'homme rationnel devant le risqué: Critique des postulates et axioms de l'ecole americaine. *Econometrica* 21, 503–546.

Bandettini, P.A., Wong, E.C., Hinks, R.S. *et al*. (1992). Time course EPI of human brain function during task activation. *Magn. Res. Med*. 25, 390–397.

Bechara, A., Damasio, H., Tranel, D., and Damasio, A. (1994). Deciding advantageously before knowing the advantageous strategy. *Science* 28, 1293–1295.

Breiter, H.C., Aharon, I., Kahneman, D. *et al*. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* 30, 619–639.

Bruni, L. and Sugden, R. (2007). The road not taken: how psychology was removed from economics, and how it might be brought back. *Economic J*. 117, 146–173.

Busino, G. (1964). Note bibliographique sur le Cours. In: V. Pareto (ed.), *Epistolario*. Rome: Accademia Nazionale dei Lincei, pp. 1165–1172.

Camerer, C., Loewenstein, G., and Prelec, D. (2005). Neuroeconomics: how neuroscience can inform economics. *J. Econ. Lit*. 43, 9–64.

Colander, D. (2007). Retrospectives: Edgeworth's hedonimeter and the quest to measure utility. *J. Econ. Persp.* 21, 215–225.

Dayan, P. and Balleine, B.W. (2002). Reward, motivation, and reinforcement learning. *Neuron* 36(2), 285–298.

Ellsberg, D. (1961). Risk, ambiguity and the savage axioms. *Q. J. Econ*. 75, 643–669.

Ferrier, D. (1878). *The Localization of Cerebral Disease*. New York, NY: G.P. Putnam and Sons.

Gallistel, C.R. (1994). Foraging for brain stimulation: toward a neurobiology of computation. *Cognition* 50, 151–170.

Gallistel, C.R. (2005). Deconstructing the law of effect. *Games Econ. Behav.* 52, 410–423.

Gigerenzer, G., Todd, P.M., and the ABC Research Group. (2000). *Simple Heuristics that Make Us Smart*. New York, NY: Oxford University Press.

Gilovich, T., Griffin, D., and Kahneman, D. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York, NY: Cambridge University Press.

Glimcher, P. (2002). Decisions, decisions, decisions: choosing a biological science of dhoice. *Neuron* 36, 323–332.

Glimcher, P. (2003). *Decisions, Uncertainty and the Brain: The Science of Neuroeconomics*. Cambridge, MA: MIT Press.

Glimcher, P.W. and Sparks, D.L. (1992). Movement selection in advance of action in the superior colliculus. *Nature* 355, 542–545.

Gold, J. and Shadlen, M. (2002). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* 36, 299–308.

Green, D.M. and Swets, J.A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: Wiley.

Gul, F. and Pesendorfer, W. (2008). The case for mindless economics. In: A. Caplin and A. Schotter (eds), *The Foundations of Positive and Normative Economics: A Handbook*. Oxford: Oxford University Press, forthcoming.

Houser, D., Bechara, A., Keane, M. *et al*. (2005). Identifying individual differences: an algorithm with application to Phineas Gage. *Games Econ. Behav.* 52, 373–385.

Houthakker, H.S. (1950). Revealed preference and the utility function. *Economic*s 17, 159–174.

Kahneman, D. and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291.

Knoch, D., Pascual-Leone, A., Meyer, K. *et al*. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832.

Kosfeld, M., Heinrichs, M., Zak, P.J. *et al*. (2005). Oxytocin increases trust in humans. *Nature* 435, 673–676.

Kwong, K.K., Bellieveau, J.W., Chesler, D.A. *et al*. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl Acad. Sci. USA* 89, 5675–5679.

Macmillan, M. (2002). *An Odd Kind of Fame: Stories of Phineas Gage*. Cambridge, MA: MIT Press.

McCabe, K., Houser, D., Ryan, L. *et al*. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl Acad. Sci. USA* 98, 11832–11835.

Montague, P.R. and Berns, G.S. (2002). Neural economics and the biological substrates of valuation. *Neuron* 36, 265–284.

Newsome, W.T., Britten, K.H., and Movshon, J.A. (1989). Neuronal correlates of a perceptual decision. *Nature* 341, 52–54.

Ogawa, S., Tank, D.W., Menon, R. *et al*. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl Acad. Sci. USA* 89, 5951–5955.

Platt, M.L. and Glimcher, P.W. (1999). Neural correlates of decision variables in parietal cortex. *Nature* 400, 233–238.

Samuelson, P.A. (1938). A note on the pure theory of consumer behavior. *Economia* 1, 61–71.

Sanfey, A.G., Rilling, J.K., Aronson, J.A. *et al*. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science* 300, 1673–1675.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron* 36, 241–263.

Shizgal, P. (1997). Neural basis of utility estimation. *Curr. Opin. Neurobiol.* 7, 198–208.

Shizgal, P. and Conover, K. (1996). On the neural computation of utility. *Curr. Direct. Psycholog. Sci.* 5, 37–43.

Smith, V. (1976). Experimental economics: induced value theory. *Am. Econ. Rev.* 66, 274–279.

Sparks, D.L. (1999). Conceptual issues related to the role of the superior colliculus in the control of gaze. *Curr. Opin. Neurobiol.* 9, 698–707.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.