

# 13 Neuroscience, Psychology, and Economic Behavior: The Emerging Field of Neuroeconomics

Paul W. Glimcher

The history of the study of judgment and decision making has been marked by an iterative tension between what are known as prescriptive and descriptive advances. Prescriptive theories, which typically have their roots in economics, seek to define efficient or optimal decision making. Descriptive empirical advances, with roots typically in psychology, then invariably suggest that these prescriptive theories do not accurately describe human behavior. The neoclassical revolution in economics during the first half of the twentieth century and the period that followed it were no exception to this general paradigm. Working from the theoretically powerful assumption that all of human behavior could be described as a rational effort to maximize a theoretical quantity known as *utility*, the neoclassical theorists largely succeeded in developing a coherent basic mathematical framework for understanding what people *should* choose. They hypothesized that there had to be some sense in which humans could be described as logically consistent, and that given this hypothesis all of the powerful tools of deductive logical mathematics could be brought to bear on the study of human decision making. This conclusion was followed, however, by a series of descriptive insights that indicated that the initial round of neoclassic theories were not consistent with human choice behavior. This meant either that humans could not be described as logically consistent in *any* sense, that the specific models developed during the neoclassical revolution were flawed, or both. The social result of this set of observations was a growing divergence between economics and psychology. In psychological circles the conviction grew that a truly logical mathematical framework for the study of decision making was not possible, while in economic circles the search for such a framework continued unabated.

One recent trend in the study of decision making may, however, reconcile this tension between the now very divergent psychological and economic approaches: a growing interest in the physical mechanisms by which human decisions are made within the human brain. Neuroeconomic scholars operating at the interface of the economic, psychological, and neurobiological domains argue that a study of the brain architecture for human decision making will reveal the actual mathematical computations that the brain performs during economic behavior. If this is true, then neurobiological studies that seek to bridge

the gap between economics and psychology may succeed in providing a methodology for reconciling prescriptive and descriptive studies of choice. These studies may produce a highly predictive and parsimonious mathematical model of individual decision making that is based on the actual computations performed by the human brain.

### **Closing the Gap between Economics and Psychology**

The revolution engendered by the advent of rational choice modeling in economics had two profound effects during the second half of the twentieth century: at a mathematical-economic level, it succeeded in defining a set of tools that could describe how an individual who wished to maximize anything (whether happiness, money, or progeny) should behave to achieve that maximization. At a behavioral-psychological level it essentially proved that humans did not reliably behave in the way predicted by the existing corpus of theory. This insight led a number of scholars at the borders of psychology and economics, perhaps most notably Herbert Simon (1947, 1983, 1997), to conclude that human decision makers could be viewed as rational utility maximizers in only a limited, or bounded, sense.<sup>1</sup> Conditions under which humans behave in accord with existing theory do occur, but there are also conditions under which humans behave in a way that contradicts existing theory. One result of this insight has been a growing conviction in some segments of the economic and psychological communities that human decision making can often be viewed as the product of two underlying processes: a bounded rational process well described by prescriptive economic theory and an irrational process which is best described empirically and which irreducibly defies formal mathematical analysis with traditional economic tools.

In response to this growing conviction, a number of scholars have recently initiated a revival of the (previously discredited) neo-Freudian neurobiological approach that dominated physiological circles in the 1950s (Freud 1923/1927; MacLean 1952). This approach suggests that two processes, the rational and irrational, are instantiated within the human brain as two anatomically discrete mechanisms. In most of these theories, like those of the 1950s, the irrational module is associated with evolutionarily ancient brain structures presumed to be irrational because of their presence in less complicated animals than ourselves. The rational module, viewed as uniquely well developed in humans, is presumed to reside in the cerebral cortex, often in frontal regions particularly highly developed in humans (McClure et al. 2004; Camerer et al. 2005). Indeed, many have suggested that irrational behavior should be uniquely attributed to limitations intrinsic to the more evolutionarily ancient portions of the brain, whereas rational behavior, when it occurs, may be viewed as the product of a conscious verbal faculty that somehow transcends this biological limitation through the use of the frontal cortex.

At the same time that this neo-Freudian approach has been revived in economic (and to a lesser extent psychological) circles, neuroscientists interested in human decision

making have begun to head in a surprisingly different direction as they seek to reconcile prescriptive and descriptive approaches. The revolution that gave birth to modern neuroscience in the early part of the twentieth century also argued that all human behavior could be conceived of as the product of two fundamentally distinct mechanisms: a sophisticated faculty that governed complex behavior, and a simpler, cruder mechanism that could produce reliable, but unavoidably simplistic (and hence implicitly irrational), behaviors (see, for example, Descartes 1664/1972; Hall 1833; Sherrington 1906). This simpler mechanism, which came to be identified with the notion of automated or reflexive responding, was widely believed to be tractable to neurophysiological analysis and formed the core of our understanding of brain function during the first half of that century.

During the last several decades, however, ongoing empirical work has begun to suggest to many neuroscientists that this view of the neural architecture is no longer tenable (Damasio 1995; LeDoux 1996; Glimcher 2003a). Biological evidence now suggests to neuroscientists a more unitary view of the neural architecture that is much more deeply rooted in evolutionary theory than this original dualistic conception. What is emerging in neuroscientific circles is the view that a surprisingly holistic (though clearly multicomponent) decision-making process governs behavior (Parker and Newsome 1998; Schall and Thompson 1999; Glimcher 2003b). The interdependent and varied inputs to this decision-making process, it is argued, have all been shaped by evolution in order to yield a unified pattern of behavior that maximizes the reproductive fitness of organisms (a rather precise and tractable definition of utility) in the environments in which they operate (Maynard Smith 1982; Stephens and Krebs 1986; Krebs and Davies 1991). Evolution makes animals fitness maximizers in a fully defined mathematical sense that has its roots in economic theory. But critically, evolution performs this role on all parts of the organism simultaneously. It yields a single whole organism, the global rationality of which is bounded not by the limits of the Freudian animal-id, but rather by the requirements of the environment within which it evolved.

This unified view stands in contrast to the neo-Freudian view, which argues that the powerful general-purpose decision-making capabilities of humans make us fundamentally different from other animals. When rationality is observed in our behavior, these scholars argue, this rationality can be attributed to a distinct and uniquely human mechanism. Quite compelling empirical data, however, argue against this conclusion. First, it now seems clear that even animals with very small brains can behave in a surprisingly rational manner under a broad range of conditions (Stephens and Krebs 1986; Krebs and Davies 1991). This argues against the idea that in order to behave rationally humans would have needed to evolve some unique faculty. Second, there is growing evidence that we share with our nearest relatives not just the ability to behave rationally but also common boundaries to our rationality (Barkow et al. 1992; Hauser 2000). If this is true, then it is both the rational and irrational that we share with our nearest relatives, again challenging the assumption that any of these aspects of behavior involve some uniquely human process. These data

argue, in essence, that we differ more in degree than in nature from our nearest living relatives.

In summary, these observations argue for three main points that will be developed below. First, a deep and successful effort to account for decision making will only be possible if scholars employ the rigorous quantitative approaches to decision making that have begun to be developed in economic circles. These models rest on mathematical logic, which is the only starting point for truly scientific studies of decision making and truly mechanistic studies of brain function. Second, although humans are unique organisms, there is growing evidence that we are far less unique in the production of decision-making behavior than many scholars at the boundary of economics and psychology suggest. For example, monkeys can play repeated mixed-strategy equilibrium games of the types Von Neumann and Morgenstern (1944) and Nash (1951) described with the same efficiency as do humans (Dorris and Glimcher 2004). Birds, to take another example, can systematically alter the shape of their utility functions to adopt risk preferences appropriate for their environments (Caraco et al. 1980). This may be the most critical point made here, because it calls into question the pervasive assumption held by many neo-Freudian economists and psychologists that our decision-making process is both a uniquely human faculty and a broadly rational faculty. Third and finally, it is absolutely critical that the economic and psychological communities recognize that neurobiological studies of decision making can be much more than efforts to locate a brain region associated with some hypothetical human faculty such as “cooperation.” Such studies are valuable starting points, but have troubled many scholars because they provide no predictive power with regard to behavior. Really useful neuroeconomic studies, from the perspective of working scientists, will have to fully describe the mechanisms by which economic computations yield observed behavior. It is an understanding of these mechanisms in that sense that will yield real predictive power in the mathematical and logical sense.

### The Neuroscience of Choice

Modern *utility theory*, the foundation of modern economics, has its origins in the theory of expected value first proposed by Pascal. He argued that the value of any course of action could be determined by multiplying the gain that could be realized from that action by the likelihood of receiving that gain. This product, which we now call *expected value*, represents the average gain or loss associated with any action. Pascal argued that when making any decision one should simply compare the expected values of the available courses of action and then select the action having the highest expected value. The most famous example of this is probably the line of reasoning from his Christian apologia, the *Pensées*, known as Pascal’s Wager. Here, Pascal (1670/1966) reasons that a belief in God is normatively rational as long as there is any uncertainty about God’s existence because the gain for believing in God is infinitely positive. Since the possible gain of eternal salvation

has infinite value, that value times any non-zero probability yields an infinite expected value, making a belief in God a rational decision.

Although Pascal and his colleagues recognized that not all human decision making could be accurately described as being guided by this concept of expected value, they argued that all rational decision making should follow this prescriptive theory (see Arnauld and Nicole 1662/1996; Pascal 1670/1966). By the early 1700s, however, it was clear that the Pascalian approach did an extremely poor job of predicting human choice behavior under conditions of significant risk.

The early psychological evidence for this conclusion arose from empirical observations about a casino game popular in St. Petersburg in the 1700s. In this game, players were asked to pay a fixed sum to participate in a single round. What they won during this round was determined by a series of coin flips. The game begins with the flip of a single coin. If that coin lands heads-up the player wins two coins. If the coin lands tails-up the coin is flipped again. If this second flip lands heads-up the player wins four coins. Otherwise, the flip repeats with the win doubling for each subsequent flip until the coin lands heads-up. Of course, the expected value for the first flip is one coin: a 50 percent chance of a heads-up times two coins. The same, however, is true for every sequential flip: for example, the a priori probability of winning in the second flip is 25 percent and the gain is four coins; of winning on the third flip 12.5 percent; and the gain is eight coins . . . From this one must conclude that the expected value of a single round of this game is infinite, although in practice players are unwilling to pay more than about forty coins per round (making this a highly unprofitable game for the casinos).

To explain this early mismatch between the prescriptive and descriptive domains the Swiss mathematician Daniel Bernoulli (1738/1954) argued for a model of rational decision making in which the likelihood of a gain was multiplied not by the objective number of coins that the chooser stood to gain, but rather by a psychological construct, now called utility, that was related to but distinct from value. His notion was that gains were represented in the psychological decision-making process by a roughly logarithmic function of value that also incorporated a representation of the chooser's wealth. Modern utility theory built on this foundation by developing a more rigorous mathematical foundation for Bernoulli's model and by explicitly recognizing that the relationship between value and utility, a relationship known as the *preference function*, is fundamentally subjective and empirical rather than being part of the prescriptively rational choice process.

Even utility theory, however, has been often challenged. Critiques of modern utility theory have tended to fall into one of two domains. The first of these classes of critiques empirically identifies failures of a specific utility-theoretic model like the ones proposed by von Neumann and Morgenstern (1944) or by Savage (1954). The second identifies behaviors for which, in principle, no truly rational model (a model that rests on basic mathematical principles) of any kind could ever account. An example of the first of these classes of failures is Kahneman and Tversky's (1979) famous observation that choosers

are more sensitive to losses than to gains. Human decision makers consider a loss of \$100 a much more negative event than they consider a gain of \$100 positive. Although this observation does challenge von Neumann and Morgenstern's (1944) model for rational choice, it does not challenge the rational framework upon which they hoped future theories would be built (although this is a point rarely made outside of economic circles). Indeed, subsequent prescriptive models that account for loss aversion, for example, have been generated by rational choice economists such as Milton Friedman and Leonard Savage (1948). The second of these classes of critique is more troubling. These critiques rest on the identification of behaviors for which no completely logical theory could account. Consider this central feature of rational choice: if I truly prefer apples to oranges then there should be no circumstances in which I can be led to voluntarily select oranges over apples in a decision-making task. Were I to prefer apples to pears and prefer pears to oranges, then I must prefer apples to oranges. The alternative, that I prefer apples to pears, pears to oranges, but oranges to apples, leads to a logical circularity (formally, a violation of the mathematical principle of transitivity) that would constitute a challenge for which no rational model could hope to account. Unfortunately, a number of these classes of behaviors have been identified by experimentalists, and it is the observation of these "preference reversals" that poses the greatest challenge for traditional economic models of choice.

An excellent example of this kind of challenge to utility theory arises in the study of choices made as a function of time, a class of behavior known as temporal discounting. In the most clear-cut example of this kind of behavior, most subjects can be shown to prefer a gain of \$22 in thirteen months over a gain of \$20 in twelve months. There is nothing irrational about this; it simply expresses a preference for the larger gain despite the additional delay. But if the same subject is asked the same question 365 days later, if he is asked whether he prefers \$20 today or \$22 in a month, changing his preference represents an inconsistency (Loewenstein and Thaler 1989). This, in a nutshell, is a critical problem for rational-choice theories because there is no way to make this pair of choices anything but logically contradictory in the mathematical sense. The contradiction arises, in a sense, because we need only to choose *when* to ask our subjects to pick in order to control their choices. Put another way, this subject's choice is inconsistent in the same way that the example of apples and oranges is inconsistent because there we can control the subjects' preference for apples simply by adjusting the order in which we present them with fruit.

In summary, then, rational-choice models from economics provide a powerful framework for understanding and modeling choice behavior—a framework that is more extensible than most scholars realize. But that framework also has clearly identifiable limits to its applicability. How, then, one might ask, should scholars interested in choice proceed? Should they discard formal models rooted in economic theory in favor of loosely defined psychological systems rooted in Freudian theory, or should they use those models, with a

clear knowledge of their limitations, as a starting point for building a new mechanistic understanding of decision making? Recent evidence suggests that the latter approach may prove the more fruitful.

Neurobiological studies conducted over the past decade have revealed that the brains of both human and nonhuman primates represent a complex variable which under many circumstances closely parallels von Neumann and Savage's notion of classical expected utility (see Platt and Glimcher 1999; Gold and Shadlen 2000; Breiter et al. 2001; Knutson et al. 2001; Paulus et al. 2001). For example, the rate at which nerve cells in the posterior parietal cortex generate action potentials is very precisely correlated with theoretically derived estimates of expected utility under many conditions (Glimcher et al. 2005). Further, some of these studies even suggest that in the final stages of the decision-making process, the neural architecture selects a course of action by mechanistically generating the response associated with the greatest activity in the posterior parietal cortex. All of these studies suggest that, despite their limitations, traditional economic theories provide tremendous descriptive power for understanding the nervous system.

### Identifying the Neurobiological Representation of Expected Utility

One of the first studies to make the suggestion that something like expected utility is actually instantiated within the nervous system was Platt and Glimcher (1999). In their experiments, trained rhesus monkeys were allowed to participate in repeated rounds of a simple lottery while the activity of nerve cells in the posterior parietal cortex was monitored. At the beginning of each round a red spot and a green spot were illuminated on a screen directly in front of the monkey. This began the lottery phase of the round, a period during which the monkey did not know whether the red or green light would be linked with a prize at the end of that round. At the end of this phase, a third light changed color to red or green, indicating which of the two initial lights had been randomly selected to yield a fruit juice reward on that particular round. The monkey then received the fruit juice if he simply made visual contact with the selected light at the end of the round. While monkeys played hundreds of rounds of this game, Platt and Glimcher systematically varied either the size of the reward associated with each light, the *value* of that light, or the relative probabilities that the red or green lights would be selected at the end of the round—the *likelihood* that each light would yield a reward.

These two variables were selected for manipulation because essentially all utility theories are based on the assumption that rational decision makers assess the desirability of any course of action by combining the value and likelihood of gain, as originally suggested by Pascal and Bernoulli. Even though in this experiment the monkeys did not need to monitor these values in order to behave efficiently, Platt and Glimcher hoped to determine whether these economic variables were encoded in the nervous system while monkeys observed these repeated lotteries, just as we might expect them to be in human players.

Platt and Glimcher found that a discrete group of nerve cells in the posterior parietal cortex encoded, separately for each light, a combination of the value and likelihood of reinforcement associated with that button during the lottery phase of each round. It appeared from this result that under these conditions the brains of their monkeys explicitly encoded something very much like the economically defined expected value or expected utility of each light in this simple lottery task.

### Game-Playing Monkeys

Dorris and Glimcher (2004) extended this finding when they examined the activity of this same brain region while a new group of rhesus monkeys engaged in a strategic conflict known as the *inspection game*. In the human version of that game, two opponents face each other, an *employer* and an *employee*. On each round of the game the employee must decide whether to *go to work*, in which case he earns a fixed wage, or whether to *shirk*, in hopes of earning his wage plus a bonus (in the human version of the game, the free time gained by shirking is itself conceived of as the bonus). The goal of the employee is simply to maximize his gain in terms of salary and bonus. The employer, on the other hand, must decide between trusting his employee to arrive for work or spending money to hire an inspector who can actually check and see whether the employee arrived for work that day. The goal of the employer is to spend as little as possible on inspections while maximizing the employee's incentive to work.

The inspection game is of particular interest to game theorists and economists because rational strategies for utility maximization during strategic conflict lead to predictable outcomes, according to the equilibrium theory originally developed by John Nash in the 1950s. Nash (1951) equilibrium theory describes how, when the cost of inspection to the employer is set high, the efficient strategy for both players converges on a solution in which the employee manages to shirk fairly often. Conversely, a low inspection cost to the employer defines a theoretical equilibrium solution in which shirk rates are low.

Dorris and Glimcher examined the behavior of both humans and monkeys during a version of the inspection game in an effort to determine whether the posterior parietal cortex really encoded something like expected utility, the theoretically defined decision variable, even under these conditions of voluntary choice. In their game, both human and monkey contestants played the role of the employee against a standardized and strategically sophisticated computer employer. Each round began with the illumination of two lights, one for working and one for shirking. At the end of each round, players selected one light and the computer employer simultaneously decided whether or not to pay for an inspection on that round. These responses were then compared by a second computer arbiter that paid both players off according to a fixed payoff matrix (paying off in juice for monkeys, real currency for humans, and virtual currency for the computer employer, as shown in figure 13.1). As in the earlier lottery task, players faced fixed conditions for a hundred or more rounds, after which the payoff matrix was changed by altering the cost of an inspection.



Inspection Game Earnings

		Computer Employer		Human Employee		Human Employee	
		Employer Action		Employer Action		Employer Action	
		Inspect	No Inspect	Inspect	No Inspect	Inspect	No Inspect
Employee Action	Work	\$1.70	\$2.00	\$1.00	\$1.00	1 oz	1 oz
	Shirk	-\$0.30	-\$2.00	\$0.00	\$2.00	0 oz	2 oz

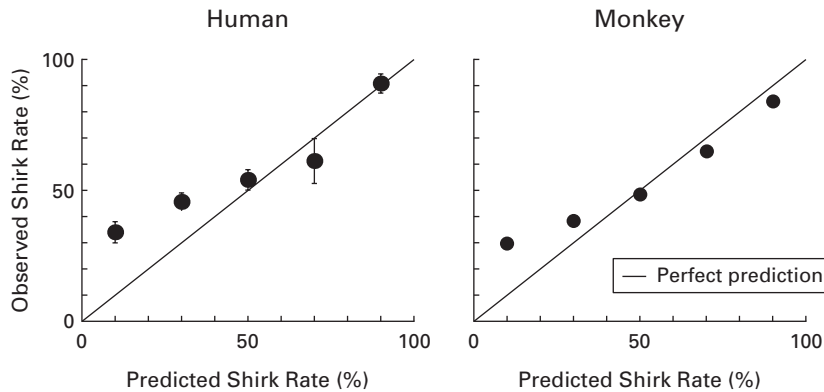
**Figure 13.1**  
 Payoffs to human and monkey employees during the inspection game (after Dorris and Glimcher 2004).

This permitted Dorris and Glimcher to examine the behavior of human and monkey players under five different sets of conditions, each of which required a slightly different strategy.

Dorris and Glimcher found that the probability a human playing the inspection game for money would choose to shirk was well predicted by the prescriptive Nash equilibrium computations whenever those computations predicted shirking rates of 40 percent or more. When, however, this particular prescriptive theory predicted shirking rates below approximately 40 percent, human subjects were observed to shirk more frequently than was predicted. This descriptive assessment of humans seemed to differ from the prescriptive assessment provided by the Nash equilibrium equations.

When Dorris and Glimcher analyzed the behavior of their monkeys, they found that the behavior of the monkeys was surprisingly similar, even essentially identical, to the behavior of their human employees. Just like humans, the monkeys seemed to precisely track the Nash equilibrium solutions and deviated from those solutions only when shirking rates of less than 40 percent were prescribed during the inspection game (figure 13.2). This was a critical advance because it allowed Dorris and Glimcher to examine the role of the posterior parietal cortex during a voluntary strategic game during which monkeys and humans seemed to employ similar, or identical, strategies.

One of Nash’s (1951) fundamental insights was that at a *mixed-strategy equilibrium*, a situation in which a strategic player should distribute her actions among two or more alternatives in an unpredictable fashion, the desirability of the two or more actions in equilibrium must be equivalent. This means that during the inspection game, the expected utilities of working and shirking must be equal, regardless of how frequently the equilibrium solution requires that the player works. The Nash approach argues, essentially, that a behavioral equilibrium occurs when the desirability of working and shirking are rendered equal by the behavior of one’s opponent, irrespective of how often that equilibrium



**Figure 13.2**

Nash Equilibrium Theory predicts human and monkey behavior equally well (after Dorris and Glimcher 2004).

requires that one work. The Nash equations themselves go a step further, defining the precise rates of working and shirking that are prescriptively rational.

Dorris and Glimcher hypothesized from the Nash approach that the desirabilities of working and shirking, rational or not, must be equivalent whenever strategic competition yields a mixed-strategy behavior in players, and thus that mixed-strategy behaviors must be associated with the equal desirability of working and shirking as represented in the nervous system. If the desirability of an action is encoded by the activity of neurons in the posterior parietal cortex not just for some categories of behavior, rational or irrational, but for behavior in general, then during strategic conflict of this type the neural activity for working and shirking should, paradoxically, always be equal. Put another way, if the economic approach is sound, then at behavioral equilibrium the desirability of working and shirking should be equivalent. If the neurobiological approach is sound, then at behavioral equilibrium the level of nerve cell activity in parietal cortex associated with working and shirking should also have been equivalent.

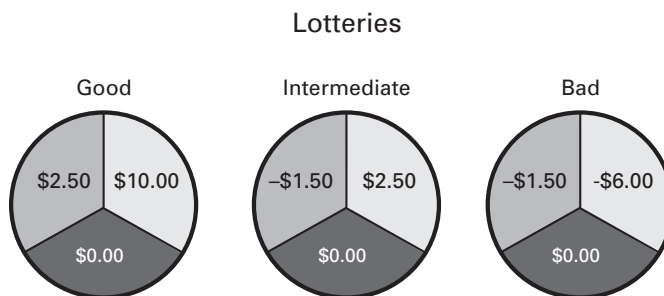
When Dorris and Glimcher examined the activity of neurons in the posterior parietal cortex while monkeys played the inspection game, they found that the posterior parietal cortex carried a signal essentially identical to the one expected. When the monkeys' behavior was well predicted by the Nash equations, neural activity was equivalent to the expected utility of economic theory. When the monkeys deviated from those prescriptive predictions, for example, by over shirking, then Dorris and Glimcher found that the activity in this area seemed to correspond to the subjective desirabilities that should have been guiding the monkeys. The neurons seemed to encode a *physiological expected utility*.

### Humans Playing Lotteries

These studies of monkeys are of importance for two reasons. First, they demonstrate the surprising similarities in the economic behavior of humans and our nearest relatives. Second, they employ highly precise brain measurement technologies that cannot be used in humans. Recently, however, the less precise brain scanning technologies that can be employed in humans have also begun to yield significant insights into the neural basis of economic behavior (McCabe et al. 2001; Montague and Berns 2002). One of the first and most compelling of these studies examined the behavior of humans during a lottery similar to the one employed by Dorris and Glimcher for the study of monkeys (Breiter et al. 2001). In that experiment, human subjects were presented, on sequential rounds, with one of three possible lotteries (see figure 13.3).

In lottery 1, the *good lottery*, they faced equal chances of winning \$10, \$2.50, or \$0. In lottery 2 they faced an equal chance of winning \$2.50, winning \$0, or losing \$1.50. In lottery 3 they faced an equal chance of winning \$0, losing \$1.50, or losing \$6.

At the beginning of each round the subjects were told which lottery they would be playing, and the average activity in many brain areas was simultaneously measured. After that measurement was complete, the lottery was actually played and the humans were then told how much real money they had earned on that round. This design was particularly interesting because of an important and well-described deviation of human behavior from prescriptive theory (Kahneman and Tversky 1979; Kahneman et al. 1982). All three of these particular lotteries present a one-third possibility of winning \$0, but they do so under different conditions. In the first lottery winning \$0 is the worst possible outcome whereas in the third lottery it is the best. Kahneman and Tversky noted that although humans rationally prefer lottery 1 to lottery 3, once they enter a lottery their perceptions of outcomes change. Once in lottery 1, winning \$0 is experienced as intensely negative while once in lottery 3 winning \$0 is experienced as positive. What Breiter and colleagues hoped to determine was whether the activity of some brain area might track both of these human responses.



**Figure 13.3**  
The three lotteries used in Breiter's experiment (after Breiter et al. 2001).

What they found was that the activity of a brain region called the subthalamic extended amygdala did behave in essentially this manner. When humans were first presented with the lottery they would face on that round, activity in this brain area was closely related to the expected utility of the lottery. After the lottery ran, however, they found that the activity of this area was a rough function of the subjective response of the human to the outcome rather than a function of the actual dollar amount won. Activity in this area was higher when the subjects won \$0 in lottery 3 than when they won \$0 in lottery 1.

Once again the neural results lead to an interesting and perhaps unexpected result. When human behavior is rational, as defined by prescriptive economic theory, we can find evidence that some brain areas encode expected utility. When, however, human behavior deviates from prescriptive theory, the brain seems to encode something more like the subjective desirability of an outcome rather than the objective economic value of that outcome.

Together, these observations raise an intriguing possibility: the neural architecture may indeed compute and represent something like the expected utility of many possible courses of action, much like that which neoclassical utility theory proposes. When choosers are efficient in the economic sense, that architecture accurately represents the objective expected utility of available choices. When economic and psychological utility differ, however, the neural architecture seems to reflect the psychological utilities that guide choice. Although it may be counterintuitive to economists to believe that subjective, or irrational, decision making reflects the principled output of highly developed neural circuits, this may simply reflect the fact that evolution shaped our neural architecture to perform efficiently under many, but not all, environmental circumstances. In some cases, inefficiencies of these types may simply arise when the most complicated cortical mechanisms inside our skulls encounter problems that they did not evolve to solve. It is these biologically based inefficiencies that therefore place boundaries on the circumstances in which we might be expected to produce economically rational behavior. The available evidence thus suggests a synthesis of modern economic and neuroscientific approaches. By biologically defining the mechanisms that compute physiological expected utility we should be able to derive a mechanistically accurate economic theory that is, by necessity, predictive.

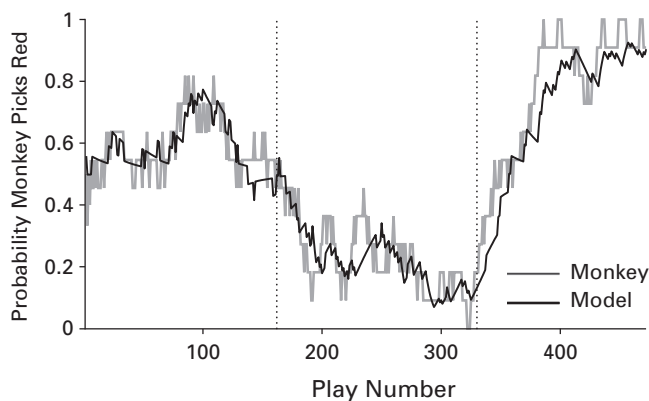
### **Using Neuroscience to Develop New Economic–Psychological Theories**

Bayer and Glimcher (2005) have been attempting to extend this approach by studying how the brain computes, or learns, the expected utilities that guide choice behavior in an effort to combine economic and psychological approaches around a neurobiological framework. They have attempted to do this by studying the activity of a group of nerve cells in the substantia nigra pars compacta that use the neurochemical dopamine to communicate with other nerve cells. These cells are widely believed to compute the difference between the

gains that a human or animal expects to receive and the gains that they actually receive (see Schultz et al. 1997; Schultz 2002), and a growing body of evidence now suggests that this is the substrate from which expected utilities, in the economic sense, may be calculated. Of particular interest from an economic point of view is the observation that this particular calculation can be shown to be prescriptively rational under some limited conditions. Of particular interest from a psychological point of view is that this particular calculation would lead to some classes of empirically observed errors under conditions where it is suboptimal. Thus, demonstrating a neural substrate that performs this calculation both when it is rational and when it is not would mechanistically unify prescriptive and descriptive studies of learning behavior.

Bayer and Glimcher therefore examined these dopamine neurons during a simple choice task in an effort to derive the precise economic equation that they compute. They then used this equation to predict the behavior of monkeys during the classic psychological matching law task of Herrnstein (see Herrnstein 1961, 1997). In that task, which was studied by Lau and Glimcher (2005), monkeys were faced with two choices reinforced on a discrete trial variable ratio schedule almost identical to the one Herrnstein studied in pigeons. On each round the monkeys could select either a red or green light placed in front of them. Before each trial began there was a fixed probability that each of the two lights would be *armed* with a reward. For example, there might be a 10 percent chance that the red light would be armed before each round and a 20 percent chance that the green light would be armed before each round, and the lights were always armed with the same amount of fruit juice. Critically, once a light was armed it remained armed until chosen by the monkey in a subsequent round.

The accompanying figure (figure 13.4) shows the free choices made by a monkey while performing as a thick gray line.



**Figure 13.4**

Predicting the choices monkeys make with a neuroeconomic model (after Lau and Glimcher 2005).

One can see that the behavior of the animal is chaotic, fluctuating from red to green. The thin black line shows the prediction of the neuroeconomic model derived from a study of the dopamine neurons. What is critical is that the model does a remarkably good job of predicting the behavior of the animal on a step-by-step basis. The model, which is neither truly prescriptive nor descriptive, is highly constrained by neurobiological observations and makes clear behavioral predictions. Of course the model is making predictions about a very simple behavior, but it seems likely at this point in time that more sophisticated models of this type will soon be developed. And it is these forthcoming models that will either validate or invalidate the promise of the developing neuroeconomic approach.

### Summary

One of the critical and persistent issues in economics has been our inability to reconcile the rational-choice model at the core of modern theory with the fact that humans are the product of a 600-million-year evolutionary lineage. We all recognize that nonhuman animals have limited mechanical and neural capacity. Fish that live in total darkness have neither eyes nor the neural architecture for vision. We all accept that even our closest living relatives, the great apes, face fundamental conceptual limitations that are probably not apparent to them. But it has long been the central premise of economic thought that humans are different from all of these other organisms. That humans rely on a more fundamentally rational neural machinery and that this machinery, which economists presume is subjectively experienced as consciousness and which they often assume is mechanistically located within the cerebral cortex, endows us with nearly perfect rationality.

In the last half century, however, a number of influential psychologists have identified conditions where humans simply do not achieve this prescriptively defined rational behavior. One conclusion that can be drawn from this is that scholars interested in understanding choice must begin to recognize that our biological-evolutionary heritage influences our actions. Many of the decisions that we make may be inefficient today because of that evolutionary history. Surprisingly, however, a group of the same economists have used this insight to argue that an accurate model of human behavior will therefore have to be two-tiered. These economists accept from classical economic theory that there is a fundamentally rational conscious decision maker within our skulls. This is, they presume, an evolutionary development unique to our species that has arisen within the very recent past. But there is also a second more ancient and mechanistic system, and when inefficient decision making occurs it can be attributed to the activity of this evolutionarily ancient mechanism.

For many neurobiologists and psychologists studying the mechanisms by which choice is accomplished, this seems to be an oddly dualist and Freudian approach to the physiology of mind. In the seventeenth century, Descartes proposed that all of human behavior could be divided into two principle classes and that each of these categories of behavior

could be viewed as the product of distinct processes. The first of those classes Descartes defined as the simple and predictable behaviors that both humans and animals could express, behaviors that predictably linked sensory stimuli with motor responses. Their simple deterministic nature suggested to him that for these behaviors the sensory-to-motor connection lay within the material body, making those simple connections amenable to physiological study. For the second class—behaviors in which no deterministic connection between sensation and action was obvious—he followed Aristotle's lead, identifying the source of these actions as the rational, but nonmaterial, soul.

Over the last several decades neurobiologists have begun to broadly reject this dualistic formulation for several reasons. First, because there seems to be no physiological evidence that such a view can be supported, and second, because it seems to fly in the face of evolutionary theory, which forms the basis of modern biology. Instead, what seems to be emerging is a much more synthetic view in which economic theory can serve as the core for a monist approach to understanding the behavior not just of simple organisms that survive in narrowly defined environments but also for understanding the most complex and generalist of extant species, *Homo sapiens*.

In sum, neuroeconomics seeks to unify the prescriptive and descriptive approaches by relating evolutionary efficiencies to underlying mechanisms. Neoclassical economics and the utility theory on which it is based provide the ultimate set of tools for describing these efficient solutions; evolutionary theory defines the field within which these mechanisms are optimized by neoclassical constraints; psychology, the empirical tools for the study of behavior; and neurobiology, the tools for elucidating those mechanisms.

Over the past decade a number of researchers in neuroscience, psychology, and economics have begun to apply this approach to the study of decision making by humans and animals. What seems to be emerging from these early studies is a basically economic view of the primate brain: the final stages of decision making seem to reflect something very much like a utility calculation. The desirability, or physiological expected utility, of all available courses of action seem to be represented in parallel, and neural maps of these physiological expected utilities seem to be the substrate upon which decisions are actually made (Glimcher 2003a).

These representations, in turn, seem to be the product of many highly coordinated brain circuits. Some of these brain circuits, such as the dopamine neurons of the substantia nigra pars compacta, are already beginning to be described. The algorithms by which these circuits compute the economic variables from which physiological expected utilities are derived are now under intensive study. Indeed, several of these mechanistic studies are even now being used to make economic predictions about the behavior of human and nonhuman primates, both when that behavior follows and when it deviates from the prescriptive neoclassical model. Studies like these seem to be elucidating the mechanisms by which economic behavior is accomplished, and a critical advantage of this approach to irrational behaviors is that once these mechanisms are understood, all behavior should

become broadly predictable. In essence, neuroeconomics argues that it is these mechanism that can serve as the logical and mathematical bridge between the prescriptive and descriptive approaches that dominate economics and psychology, respectively.

As early as 1898 the economist Thorstein Veblen made this point in an essay entitled “Why Is Economics Not an Evolutionary Science?” He suggested that in order to understand the economic behavior of humans one would have to understand the mechanisms by which those behaviors were produced. More recently the biologist E. O. Wilson (1998) has made a similar point. Arguing that a fusion of the social and natural sciences is both inevitable and desirable, Wilson has suggested that this fusion will begin with a widespread recognition that economics and biology are two disciplines addressing a single subject matter. Ultimately, economics and psychology are biological sciences. They are the study of how humans behave. That behavior is inescapably a biological process. Truly understanding how and why humans make the choices that they do will undoubtedly require a neuroeconomic science.

### Note

1. The term “utility” has often been the subject of profound misconceptions. When von Neumann and Morgenstern (1944), and Savage (1954) defined utility they meant it to be the internal experience associated with any possible event in the outside world that guided decision making. Love, social status, and of course money were all meant to be the subjects of utility theories. Their goal was simply to describe how choosers, given an individual-specific mapping between events in the world and utility, should have to maximize that utility. Subsequent theorists have tended to focus on the maximization of monetary wealth because of the importance of wealth to the economy. One unfortunate side effect of this focus, however, has been a misunderstanding of utility theory and its goals. Utility theory is not about maximizing money. The observation that humans care about quality, will forego money to protect their children, or make different decisions as they age, poses no particular problems for utility theory. Utility theory simply asks whether or not there is a conceptual framework under which the mathematical tools of deductive logic can be applied to the study of choice.

### References

- Arnould A, Nicole P (1662/1996) *Logic, or the art of thinking* (Buroker JV, ed) Cambridge: Cambridge University Press.
- Barkow J, Cosmides L, Tooby J, eds (1992) *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.
- Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47: 129–141.
- Bernoulli D (1738/1954) Exposition of a new theory on the measurement of risk. *Econometrica* 22: 23–36.
- Breiter HC, Aharon I, Kahneman D, Dale A, Shizgal P (2001) Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* 30: 619–639.
- Camerer C, Loewenstein G, Prelec D (2005) Neuroeconomics: How neuroscience can inform economics. *J Econ Lit* 43: 9–64.
- Caraco T, Martindale S, Whittam TS (1980) An empirical demonstration of risk-sensitive foraging preferences. *Anim Behav* 28: 820–830.
- Damasio AR (1995) *Descartes’s error: Emotion, reason and the human brain*. London: Pan Macmillan.



- Descartes R (1664/1972) *L'Homme* (Hall TS, trans). Cambridge, MA: Harvard University Press.
- Dorris MC, Glimcher PW (2004) Activity in posterior parietal cortex is correlated with the subjective desirability of an action. *Neuron* 44: 365–378.
- Freud S (1923/1927) *The ego and the id*. London: Hogarth Press.
- Friedman M, Savage L (1948) The utility analysis of choices involving risk. *J Polit Econ* 56: 279–304.
- Glimcher PW (2003a) *Decisions, uncertainty and the brain: The science of neuroeconomics*. Cambridge, MA: MIT Press.
- Glimcher PW (2003b) Neural correlates of primate decision making. *Annu Rev Neurosci* 25: 133–179.
- Glimcher PW, Dorris MC, Bayer HM (2005) Physiological utility theory and the neuroeconomics of choice. *Games Econ Behav* 52: 213–256.
- Gold JI, Shadlen MN (2000) Representation of a perceptual decision in developing oculomotor commands. *Nature* 404: 390–394.
- Hall M (1833) On the reflex function of the medulla oblongata and medulla spinalis. *Philos T Roy Soc* 123: 635–665.
- Hauser M (2000) *Wild minds: What animals really think*. New York: Henry Holt.
- Herrnstein RJ (1961) Relative and absolute strength of response as a function of frequency of reinforcement. *J Exp Anal Behav* 4: 267–272.
- Herrnstein RJ (1997) *The matching law* (Rachlin H, Laibson DI, eds). Cambridge, MA: Harvard University Press.
- Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263–291.
- Kahneman D, Slovic P, Tversky A (1982) *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Knutson B, Adams CM, Fong GW, Hommer D (2001) Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci* 21: RC159.
- Krebs JR, Davies NB, eds (1991) *Behavioural ecology*. Oxford: Blackwell Scientific Publications.
- Lau B, Glimcher PW (2005) Dynamic response-by-response models of matching behavior in rhesus monkeys. *J Exp Anal Behav* 84: 555–579.
- LeDoux J (1996) *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster.
- Loewenstein G, Thaler R (1989) Anomalies: Intertemporal choice. *Journal Econ Perspect* 3: 181–193.
- MacLean PD (1952) Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). *Electroen Clin Neuro Suppl* 4: 407–418.
- Maynard Smith J (1982) *Evolution and the theory of games*. Cambridge: Cambridge University Press.
- McCabe K, Houser D, Ryan L, Smith V, Trouard T (2001) A functional imaging study of cooperation in two-person reciprocal exchange. *P Natl Acad Sci USA* 98: 11832–11835.
- McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate neural systems value immediate and delayed monetary rewards. *Science* 306: 503–507.
- Montague PR, Berns GS (2002) Neural economics and the biological substrates of valuation. *Neuron* 36: 265–284.
- Nash JF (1951) Non-cooperative games. *Ann Math* 54: 286–295.
- Parker AJ, Newsome WT (1998) Sense and the single neuron: Probing the physiology of perception. *Annu Rev Neurosci* 21: 227–77.
- Pascal B (1670/1966) *Pensées* (Kraillsheimer AJ, trans). London: Penguin Books.
- Paulus MP, Hozack N, Zauscher B, McDowell JE, Frank L, Brown GG, Braff DL (2001) Prefrontal, parietal, and temporal cortex networks underlie decision making in the presence of uncertainty. *Neuroimage* 13: 91–100.
- Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400: 233–238.

- Savage L (1954) *The foundations of statistics*. New York: Wiley.
- Schall JD, Thompson KG (1999) Neural selection and control of visually guided eye movements. *Annu Rev Neurosci* 22: 241–59.
- Schultz W (2002) Getting formal with dopamine and reward. *Neuron* 36: 241–263.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275: 1593–1599.
- Sherrington CS (1906) *The integrative action of the nervous system*. New York: Charles Scribner's Sons.
- Simon HA (1947) *Administrative behavior*. New York: Free Press.
- Simon HA (1983) *Reason in human affairs*. Palo Alto: Stanford University Press.
- Simon HA (1997) *Models of bounded rationality: Empirically grounded economic reason*. Cambridge, MA: MIT Press.
- Stephens DW, Krebs JR (1986) *Foraging theory*. Princeton: Princeton University Press.
- Veblen T (1898) Why is economics not an evolutionary science? *Q J Econ* 12: 373–397.
- Von Neumann JV, Morgenstern O (1944) *Theory of games and economic behavior*. Princeton: Princeton University Press.
- Wilson EO (1998) *Consilience*. New York: Knopf.