



## Neuroeconomics: The Consilience of Brain and Decision

Paul W. Glimcher and Aldo Rustichini

*Science* **306**, 447 (2004);

DOI: 10.1126/science.1102566

---

*This copy is for your personal, non-commercial use only.*

---

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of July 11, 2013):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/306/5695/447.full.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/306/5695/447.full.html#related>

This article **cites 37 articles**, 13 of which can be accessed free:

<http://www.sciencemag.org/content/306/5695/447.full.html#ref-list-1>

This article has been **cited by** 132 article(s) on the ISI Web of Science

This article has been **cited by** 42 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/306/5695/447.full.html#related-urls>

This article appears in the following **subject collections**:

Psychology

<http://www.sciencemag.org/cgi/collection/psychology>

4. S. Ito, V. Stuphorn, J. W. Brown, J. D. Schall, *Science* **302**, 120 (2003).
5. K. Shima, J. Tanji, *Science* **282**, 1335 (1998).
6. J. O'Doherty, M. L. Ringelbach, E. T. Rolls, J. Hornak, C. Andrews, *Nature Neurosci.* **4**, 95 (2001).
7. M. Ullsperger, D. Y. von Cramon, *J. Neurosci.* **23**, 4308 (2003).
8. C. B. Holroyd, M. G. H. Coles, *Psychol. Rev.* **109**, 679 (2002).
9. C. B. Holroyd, J. T. Larsen, J. D. Cohen, *Psychophysiology* **41**, 245 (2004).
10. H. Gemba, K. Sasaki, V. B. Brooks, *Neurosci. Lett.* **70**, 223 (1986).
11. M. Ullsperger, D. Y. Von Cramon, *Cortex*, in press.
12. N. Picard, P. L. Strick, *Cereb. Cortex* **6**, 342 (1996).
13. C. B. Holroyd, S. Nieuwenhuis, R. B. Mars, M. G. H. Coles, in *Cognitive Neuroscience of Attention*, M. I. Posner, Ed. (Guilford, New York, in press).
14. W. J. Gehring, B. Goss, M. G. H. Coles, D. E. Meyer, E. Donchin, *Psychol. Sci.* **4**, 385 (1993).
15. M. Falkenstein, J. Hoormann, S. Christ, J. Hohnsbein, *Biol. Psychol.* **51**, 87 (2000).
16. W. Schultz, *Neuron* **36**, 241 (2002).
17. C. B. Holroyd et al., *Nature Neurosci.* **7**, 497 (2004).
18. M. M. Botvinick, T. S. Braver, D. M. Barch, C. S. Carter, J. D. Cohen, *Psychol. Rev.* **108**, 624 (2001).
19. N. Yeung, M. M. Botvinick, J. D. Cohen, *Psychol. Rev.*, in press.
20. Materials and methods are available as supporting material on Science Online.
21. R. Hester, C. Fassbender, H. Garavan, *Cereb. Cortex* **14**, 986 (2004).
22. The majority of activations fall into the border zone between areas 8, 6, and 32, with some extension into area 24. Recent research in nonhuman primates seems to suggest a functional-anatomical dissociation of regions subserving pre-response conflict monitoring from structures sensitive to errors and omission of reward (1, 4). Although in humans this view is still under debate (11, 13, 21), the present meta-analysis does not provide unequivocal evidence for or against such a dissociation. Activations related to pre-response conflict and uncertainty occur more often in area 8 and less often in area 24 than do signal increases associated with errors and negative feedback (area 8, 32.5% versus 9.7%; area 24, 7.5% versus 25.8%), supporting the dissociation view. However, both groups of activations cluster primarily in area 32 (pre-response, 42.5%; error, 41.9%), suggesting that pre- as well as post-response monitoring processes share at least one underlying structure. It seems that the currently available spatial resolution in fMRI, in conjunction with anatomical variability and differences in scanning and preprocessing methods between studies, limit the ability to resolve this debate about a possible dissociation in the range of 10 mm or less.
23. T. Paus, *Nature Rev. Neurosci.* **2**, 417 (2001).
24. H. D. Critchley et al., *Brain* **126**, 2139 (2003).
25. E. K. Miller, J. D. Cohen, *Annu. Rev. Neurosci.* **24**, 167 (2001).
26. A. R. Aron, T. W. Robbins, R. A. Poldrack, *Trends Cogn. Sci.* **8**, 170 (2004).
27. D. Badre, A. D. Wagner, *Neuron* **41**, 473 (2004).
28. S. A. Bunge, K. N. Ochsner, J. E. Desmond, G. H. Glover, J. D. E. Gabrieli, *Brain* **124**, 2074 (2001).
29. M. Brass, D. Y. von Cramon, *J. Cogn. Neurosci.* **16**, 609 (2004).
30. K. R. Ridderinkhof et al., *Science* **298**, 2209 (2002).
31. K. R. Ridderinkhof, S. Nieuwenhuis, T. R. Bashore, *Neurosci. Lett.* **348**, 1 (2003).
32. J. G. Kerns et al., *Science* **303**, 1023 (2004).
33. H. Garavan, T. J. Ross, K. Murphy, R. A. Roche, E. A. Stein, *Neuroimage* **17**, 1820 (2002).
34. G. Bush et al., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 523 (2002).
35. J. O'Doherty, H. Critchley, R. Deichmann, R. J. Dolan, *J. Neurosci.* **23**, 7931 (2003).
36. R. Coles, L. Clark, A. M. Owen, T. W. Robbins, *J. Neurosci.* **22**, 4563 (2002).
37. J. F. Bates, P. S. Goldman-Rakic, *J. Comp. Neurol.* **336**, 211 (1993).
38. M. Petrides, D. N. Pandya, *Eur. J. Neurosci.* **11**, 1011 (1999).
39. L. Koski, T. Paus, *Exp. Brain Res.* **133**, 55 (2000).
40. K. Fiehler, M. Ullsperger, D. Y. von Cramon, *Eur. J. Neurosci.* **19**, 3081 (2004).
41. W. J. Gehring, R. T. Knight, *Nature Neurosci.* **3**, 516 (2000).
42. M. I. Posner, G. J. DiGirolamo, in *The Attentive Brain*, R. Parasuraman, Ed. (MIT Press, Cambridge, MA, 1998), pp. 401–423.
43. K. Matsumoto, K. Tanaka, *Science* **303**, 969 (2004).
44. K. Matsumoto, K. Tanaka, *Curr. Opin. Neurobiol.* **14**, 178 (2004).
45. This research was supported by a TALENT grant (E.A.C.) and a VENI grant (S.N.) of the Netherlands Organization for Scientific Research and by the Priority Program Executive Functions of the German Research Foundation (M.U.). Helpful comments by S. Bunge are gratefully acknowledged.

**Supporting Online Material**  
[www.sciencemag.org/cgi/content/full/306/5695/443/DC1](http://www.sciencemag.org/cgi/content/full/306/5695/443/DC1)  
 Materials and Methods  
 Table S1  
 References

## REVIEW

# Neuroeconomics: The Consilience of Brain and Decision

Paul W. Glimcher<sup>1\*</sup> and Aldo Rustichini<sup>2</sup>

Economics, psychology, and neuroscience are converging today into a single, unified discipline with the ultimate aim of providing a single, general theory of human behavior. This is the emerging field of neuroeconomics in which consilience, the accordance of two or more inductions drawn from different groups of phenomena, seems to be operating. Economists and psychologists are providing rich conceptual tools for understanding and modeling behavior, while neurobiologists provide tools for the study of mechanism. The goal of this discipline is thus to understand the processes that connect sensation and action by revealing the neurobiological mechanisms by which decisions are made. This review describes recent developments in neuroeconomics from both behavioral and biological perspectives.

*The full understanding of utility will come from biology and psychology by reduction to the elements of human behavior followed by a bottom-up synthesis, not from the social sciences by top-down inference and guesswork based on intuitive knowledge. It is in biology and psychology that economists and social scientists will find the*

*premises needed to fashion more predictive models, just as it was in physics and chemistry that researchers found the premises that upgraded biology.* (p. 206) (1)

Consider the famous St. Petersburg paradox (2). Which of the following would you prefer, \$40 or a lottery ticket that pays according to the outcomes of one or more fair coin tosses: heads you get \$2 and the game ends, tails you get another toss and the game repeats, but now if the second toss lands heads up you get \$4, and so on. If the *n*th toss is the first to land heads up, you get

2<sup>*n*</sup> dollars. The game continues, however long it takes, until the coin lands heads up. We can assess the average objective, or expected, value of this lottery by multiplying the probability of a win on each flip by the amount of that win:

$$\begin{aligned} \text{Expected value} &= (0.5 \times 2) + (0.25 \times 4) + \\ &\quad (0.125 \times 8) \dots \\ &= 1 + 1 + 1 + \dots \end{aligned}$$

This simple calculation reveals that the expected value of the lottery is infinite even though the average person is willing to pay less than \$40 to play it. How could this be?

For an economist, any useful explanation must begin with a set of assumptions that renders behavior formally tractable to coherent theoretical and mathematical analysis. Economists therefore explain this behavior by assuming that the desirability of money does not increase linearly, but rather grows more and more slowly as the total amount at stake increases. For example, the desirability of a given amount might be a power function

<sup>1</sup>Center for Neural Science, New York University, New York, NY 10003, USA. <sup>2</sup>Department of Economics, University of Minnesota, Minneapolis, MN 55455, USA.

\*To whom correspondence should be addressed. E-mail: [glimcher@cns.nyu.edu](mailto:glimcher@cns.nyu.edu)

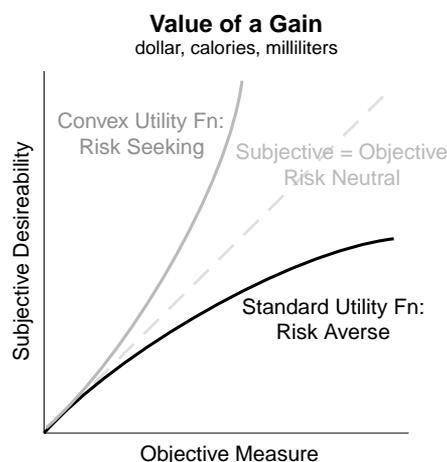
of that amount, as shown by the black line in Fig. 1. A decision-maker for whom the subjective value, or utility, of money grew in this fashion would then determine the desirability, or expected utility, of the St. Petersburg lottery by multiplying the probability of a win on each flip by the utility of the amount won on that flip, and thus he might well be willing to pay less than \$40 to play this game.

From the point of view of a psychologist attempting to understand and explain this same phenomenon, it is the nature of risk aversion and the psychological mechanisms that this set of preferences reveals that become the subject of explanatory study. The psychological mechanism that accounts for risk aversion in human subjects, for example, has been shown to be more sensitive to monetary losses than to monetary gains. Further, psychologists have suggested that subjective utilities are computed with regard to somewhat arbitrary and idiosyncratic monetary reference points, or frames, set by yet other psychological processes (3). Psychologists use observations like these to argue that human choosers are endowed with a particularly strong fear of losses and that they weigh the merits of all possible gains and losses relative to a psychological benchmark: The psychological approach seeks empirically to describe minimally complex behavioral tendencies, modules, or heuristics that can account for the actions of human choosers.

A traditional neurobiological perspective uses yet another approach: A hungry bird is shown a tray that contains five millet seeds and repeatedly permitted to fly to the tray and eat the seeds. At a neurobiological level, the study of this behavior begins with the assumption that the visual stimulus of the five seeds must somehow propagate through the sensory system of the animal to trigger activation in orienting circuits that move the bird to the seeds. Next, the same bird is permitted to fly to a second tray covered by a piece of paper. When the bird displaces the cover, half of the time it reveals 12 seeds and half of the time it reveals nothing. Mechanistically, the visual stimulus must again trigger an orienting response, and presumably in this case the strength with which visual signals connect synaptically to the orienting circuits reflects both the number of seeds that the bird earns and the likelihood that seeds will be found under the paper. Lastly, both trays are presented, and the bird is observed to fly toward the tray that may contain 12 millet seeds. A standard neurobiological explanation (4, 5) presumes that under these circumstances the two different behavioral circuits compete. In this case the synapses that elicit an orienting response to the covered tray are stronger and thus control

behavior. The neurobiological explanation specifies the minimal neural circuitry required to account for the observed behavior of the bird.

What is striking about explanations of choice behavior by economists, psychologists, and neurobiologists is the different levels at which they operate. The economic approach attempts to describe globally all choice behavior with a single logically consistent formalism. The psychological approach examines the ways in which subjective and objective estimates of value differ and posits psychological modules that can account for these observed behavioral preferences. The neurobiological explanation starts with the simplest possible neural



**Fig. 1.** Bernoulli's notion of subjective value or utility. The black line plots the typical relationship between objective and subjective valuations of an action. As the objective value of a gain increases, the subjective desirability, or utility, grows more slowly. Bernoulli demonstrated that this relationship could account for the observation that humans are typically risk-averse. The solid gray line plots a condition in which subjective value grows more quickly than objective value, a preference structure that would yield risk-seeking behavior.

circuitry that can account for the simplest measurable elements of behavior. It seems obvious that these different levels of explanation should be linked, but how can such a linkage be accomplished? We argue that a unified explanation of decision-making is not only possible but has recently begun and that, when the linkage between these three levels of explanation has matured, a new, more powerful decision science rooted in a neuroeconomic approach will have been developed.

A second claim we will make is that once this reconstruction of decision science is completed, many of the most puzzling aspects of human behavior, aspects that economic theory, psychological analysis, or neurobiological deconstruction have failed to

explain, will become formally and mechanistically explicable. The claim is, in essence, that a decision science that simultaneously engaged all three approaches would be more heavily constrained and at the same time would have much greater explanatory power than do any of these three approaches operating alone. We will see examples of how this synthetic approach would operate in principle and early attempts at synthetic solutions below.

This reconstruction of the study of decision is also going to be the appropriate basis for a more ambitious theory that explains not just how we make decisions but why. That such an explanation is necessary and possible is indicated by the fact that fundamental features of decision making are common to many species. For example, risk aversion as shown by the St. Petersburg paradox has been described in many species. Studies of birds making choices in risky environments produce a behavior best described by a utility function (Fig. 1) (6, 7). We know that humans and birds deviated from a common reptilian ancestor at least 200 million years ago, but this basic function for choice has remained essentially unchanged. Such commonalities make a clear suggestion: A utility function of this type probably is an efficient and evolved feature of vertebrate choice. For example, Robson (8) provides a justification of why a utility function might be an evolutionary optimal response to changing environments. Just as information theory was used by Barlow (9) to explain why animals as diverse as horseshoe crabs and cats use similar encoding schemes in their visual systems, an economic theory that relates utility to Darwinian fitness must serve as an overarching tool for understanding vertebrate choice behavior.

### Linking the Decision Sciences

*Subjective desirability.* The central concept in modern economic theory is the notion of subjective utility: Preferences must be described as subjective properties of the chooser. Surprisingly, the notion that preferences are represented in the nervous system, that these preferences are subjective, and that they guide the production of action has only recently entered the neurobiological mainstream. We believe that this has been a critical flaw in neurobiological studies, because it is essential that economics, psychology, and neuroscience acknowledge a common phenomenological base to achieve a reductive unification of the decision sciences. The concepts that guide the behavioral study of decision-making must also guide the mechanistic study of that process.

In part, this preference-free approach may have arisen from neurobiology's roots in the stimulus-response physiology of the

early twentieth century (10). Working from Descartes' philosophy, Sherrington (11) proposed that physiologists should work to link stimulus and response directly through what Pavlov (12) would later call a "definite nervous path." Scientists like Sherrington and Pavlov proposed that it was the role of neuroscience to chart these stimulus-response connections through the nervous system.

A critical step beyond this initial framework was a recent effort to explain more complicated behaviors and to focus on actions for which deterministic sensory-to-movement mapping approaches were insufficient. Newsome and his colleagues (13, 14) made that step in the late 1980s when they examined perceptual decision-making by monkeys viewing ambiguous sensory stimuli. In those experiments, monkeys stared at a display of chaotically moving spots of light. On training trials, a subset of the spots moved coherently in a single direction, whereas the remaining spots moved randomly (15). The direction of this coherent motion indicated which of two possible saccadic eye movements would yield a fruit juice reward, and at the end of each trial animals were free to make a saccade. If they made the correct movement, they then received the reward. On a critical subset of trials, however, monkeys viewed displays in which none of the dots moved in a single coherent direction, and thus the display provided no information from which the location of the rewarded eye movement could be deduced. Under these conditions, Newsome and his colleagues found that the firing rates of single neurons in the middle temporal visual area (area MT) were still correlated with the behavior of the animals, even when that behavior could not be predicted from the properties of the stimulus. Newsome and his colleague Shadlen's subsequent studies revealed the basic neurobiological substrate for perceptual decision-making and showed convincingly that this circuit could not be modeled simply as a single "definite nervous path" from stimulus to response (16).

This work, in turn, accelerated studies of the posterior parietal cortex, an area interposed between many of the sensory circuits and motor circuits of the primate brain, which appeared to play a critical role in the perceptual decision-making Newsome has studied (17, 18). Platt and Glimcher (19) made an important advance when they extended Newsome's approach by proposing that posterior parietal cortex might play a role in decision-making in an economic

sense and that it might encode the desirabilities of making particular movements.

In Platt and Glimcher's 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, two yellow spots were illuminated on a screen, one to the left and one to the right of where the monkey was looking. This began the lottery phase of the round, a period during which the monkey did not know whether the left or right light would be offered as 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 received the fruit juice if he oriented to 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 relative probabilities that the left or right lights would be selected at

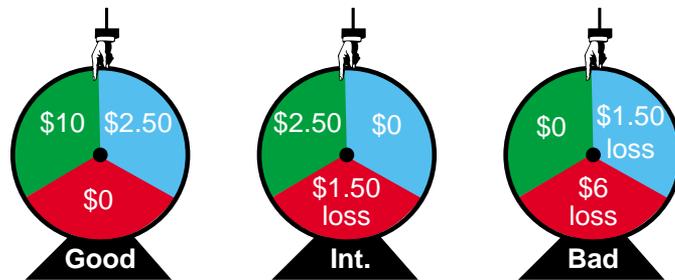


Fig. 2. The three lotteries used in the Breiter and colleagues experiment.

the end of each round or the size of the reward associated with each. These two variables were selected because economic theories assess desirability by combining the value and likelihood of gain in some subjective manner. Platt and Glimcher found that some parietal neurons did indeed encode the value and likelihood of reinforcement during the lottery phase of each round. Under these conditions, the brains of the monkeys explicitly encoded something very much like the economically defined expected value or expected utility of each light in this simple lottery task.

Subsequent studies of human decision-making using functional magnetic resonance imaging (fMRI) have yielded similar conclusions. Knutson and colleagues (20) have shown, for example, that activity in the human striatum is correlated with the magnitude of the monetary reward subjects earn during lotteries, and Paulus and colleagues (21) have shown a similar result in the human posterior parietal cortex. In a particularly interesting study, Breiter and col-

leagues (22) (Fig. 2) presented human subjects on sequential rounds with one of three possible lotteries. In lottery one (the good lottery), they faced equal chances of winning \$10, \$2.50, or \$0. In lottery two (the intermediate lottery), they faced equal chances of winning \$2.50, winning \$0, or losing \$1.50. In lottery three (the bad lottery), they faced equal chances 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. Importantly, all three of these lotteries present a one-third possibility of winning \$0, but they do so under different conditions. In the good lottery, winning \$0 is the worst possible outcome, whereas in the bad lottery it is the best. The psychologists Kahneman and Tversky (23) have shown that, when a human participates in this good lottery, they find winning \$0 to be an intensely negative outcome whereas when a human participates in the bad lottery, they find winning \$0 to be a positive outcome; subjective utilities are computed with regard to a reference frame. Breiter and colleagues found that the activity of the sublenticular extended amygdala encoded the desirability of each lottery, taking into account this behaviorally described framing effect.

Other recent neurobiological studies have revealed yet other neurally encoded variables that include the log likelihood that a given eye movement will result in a reward (24, 25), the very closely related integral of perceptual signals indicating which saccade will be rewarded in the Newsome task (26, 27), the average rate at which a saccade has been rewarded in the recent past (28), the instantaneous likelihood, or hazard, that a reinforced saccade will be instructed (29), and combinations of these variables (30).

*Strategic thinking.* All of these results suggest that classical utility theory can be used as a central concept for the study of choice in economics, psychology, and neuroscience. In the middle of the twentieth century, however, economists pushed utility theory beyond this boundary, enhancing it to include the study of the strategic interactions which arise when decision makers confront intelligent opponents. Extending the concept of subjective utility, VonNeumann and Morgenstern (31) and Nash (32, 33) developed a formal utility-based economic approach in the theory of

games. Recently, Lee and his colleagues (34) and Dorris and Glimcher (35) have begun to link the neurobiological corpus to this literature by examining the activity of single neurons in awake-behaving monkeys engaged in decision-making during strategic conflicts. In Dorris and Glimcher's study, 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. The goal of the employee is simply to maximize his gains 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 this strategic conflict lead to predictable outcomes according to an equilibrium theory originally developed by John Nash in the 1950s. Nash (36) 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. One of the fundamental insights this formal analysis reveals is that at a mixed strategy equilibrium, a situation in which a rational player should distribute his actions amongst two or more alternatives in an unpredictable fashion, the desirability of the two or more actions in equilibrium must be equivalent. 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 requires that one work. When Dorris and Glimcher examined the activity of neurons in the posterior parietal cortex of monkeys playing the inspection game, they found neurons that carried a signal that behaved like relative expected utility. When the monkey's behavior was well predicted by the Nash equations, neural activity was fixed at a single level irrespective of the frequency with which the monkey

chose to make a particular response, even though these same neurons were strongly modulated by changes in the value of actions during lottery tasks.

Research on human-human strategic interactions that are well described by classical game theory are also now under way in a number of laboratories (37). Like the earlier fMRI studies of simple decision-making tasks, these experiments are also beginning to shape the common ground between economics, psychology, and neuroscience. Taken together, these findings suggest that at least under some circumstances decisions may actually be made in the primate neuro-architecture in a manner long suspected by economists and now being actively analyzed by psychologists and neuroscientists: Neural circuits may compute and represent the desirability of making a response. Economics, psychology, and neuroscience do seem to be

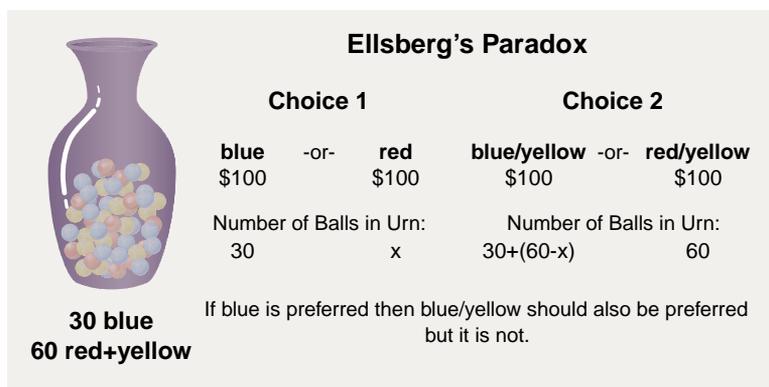


Fig. 3. Ellsberg's paradox. If blue is preferred in choice one, then blue/yellow should logically be preferred in choice two. Surprisingly, this is rarely the case.

converging around a common conceptual framework. All three disciplines are beginning to acknowledge that decision-making involves the representation of subjective desirabilities. The challenge that this convergence around a single concept poses, however, is to leverage the intersection of these three disciplines to explain choice behavior that cannot be described with the common framework of utility theory; these are classes of behaviors which have stymied traditional economics and which have lain far beyond the reach of traditional neuroscience. If it is to be of value, the goal of a unified decision science will have to be to use all three sets of approaches simultaneously to gain traction in this new territory.

### Beyond Classical Concepts

*Choice under risk.* As we have seen, the introduction of the concept of expected utility solved the St. Petersburg's puzzle and formed the core of neoclassical economics. Sub-

sequent puzzles and paradoxes, however, have plagued this solution. In Ellsberg's (1961) paradox (38) (Fig. 3) you are presented with an urn, and you are told that it contains 90 balls. Of these, 30 are blue, and 60 are either red or yellow; any proportion is possible. You are then offered a choice between a lottery that pays \$100 if a blue ball is drawn (a 1/3 probability) and one that pays \$100 if a red ball is drawn. The probability of a red draw is unspecified or ambiguous: It is a choice between an event with a known probability and an event with an unknown probability. Under these circumstances, people typically choose the first lottery, which wins if a blue ball is drawn. According to expected utility theory they could only do so if they believe that there are fewer than 30 red balls in the urn or, equivalently, that there are more than 30 yellow balls. Then (before any balls are actually drawn, but with the same urn standing in front of you) you

are asked to choose again, this time between a lottery that pays \$100 on either blue or yellow and one that pays \$100 on either red or yellow. Now the likelihood of winning is clear in the second case (a 2/3 probability of winning \$100) but unclear in the first case (a probability between 1/3 and 1). People this time typically choose the second lottery. The first lottery seems less attractive, because there might be too few yellow balls. Is there anything wrong with this behavior? If expected utility

theory is correct, then there certainly is: You cannot think that there are too few and too many yellow balls in the urn at the same time.

The Ellsberg paradox is just one of many demonstrations presented in the last half of the twentieth century that were considered formal falsifications of expected utility theory. An even earlier example is Allais' paradox (39), based on the idea that a certain outcome may be perceived as more desirable, in a qualitatively different way, than any random outcome, even if very likely (40). These examples proved that expected utility theory as originally proposed could not be globally correct; at best it could only predict choices under some circumstances. This has led economists and social psychologists both to attempt modifications to expected utility theory and to replace it outright. Both the modifications and replacements have provided important and economically powerful insights into choice behavior but have not yet provided a global theory of

choice that can truly replace expected utility theory.

The emerging discipline of neuroeconomics offers a new strategy both for testing existing models of all types and for developing new models with empirical techniques. If we succeed in understanding mechanistically how choices that violate expected utility theory are made at a neural level, then a new global theory of choice will be developed. To that end, a number of laboratories are now beginning to reexamine the conditions under which expected utility theory fails.

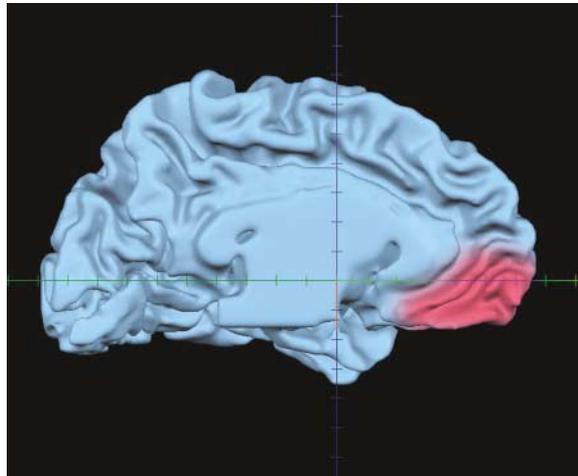
The reason that expected utility theory fails under some conditions may be that choosers use more than one evaluative mechanism at a neurobiological level (41). For example, in Dickhaut *et al.* (42), the processes involved when a certain outcome is one of the options are different from those involved when only random outcomes are at stake, providing an explanation of the Allais' paradox cited above.

Under many conditions these mechanisms may work together to yield choices similar to those predicted by expected utility theory but may produce odd results when used in isolation, in novel combinations, or in situations for which they are ill suited. Recent work by Damasio and colleagues [for example, (43)] on the class of behavioral paradoxes from which the Ellsberg example is drawn seem to support this conclusion. These studies suggest that an ambiguity-sensitive mechanism associated with the expression of emotion may reside, at least in part, in the ventromedial prefrontal cortex (VMPFC) (Fig. 4) and may be responsible for choice under some but not all conditions. These researchers and others have shown that patients with damage to this area have an impaired ability to make some classes of decisions and have difficulties planning their work and choosing friends. Further, the actions these individuals do elect to pursue often lead to financial as well as personal losses. Yet despite these specific failures, patients with damage to the VMPFC show normal performance on multiple-choice tests of intelligence.

These observations and others like them have led Damasio to propose that the inability of patients with VMPFC lesions to make advantageous decisions under some circumstances is caused by damage to an emotional mechanism that stores and signals the value of future consequences of an action, the somatic marker hypothesis. The hypothesis proposes that, because they lack this emotional mechanism, the patients

must rely on other brain mechanisms that achieve a different analysis of the numerous and often conflicting options involving both immediate and future consequences. This other mechanism, operating alone, is hypothesized to produce decisions that are less efficient and slower than those produced by a normal, intact, system.

The importance of the emotion-related VMPFC for regular decision-making has been confirmed by experiments where subjects were asked to make choices among a group of alternatives that carry a monetary reward (typically by selecting one card at a time from four different decks of cards), but for which the probability of reward is unspecified (44). This is precisely the ambiguous situation that produces Ellsberg's paradox. Under these conditions, patients with VMPFC lesions seem to lack



**Fig. 4.** Medial view of the left half of a human brain, with the front of the brain on the right side of the image. The human ventromedial prefrontal cortex is shown in red.

an aversion to ambiguity or losses that normal subjects have, an aversion that may be quite advantageous under many conditions. Further support for this hypothesis comes from brain imaging studies. For example, O'Doherty *et al.* (45) have shown that the VMPFC is relatively more active when human subjects are actively learning about the availability of rewards and punishments during one of these ambiguous choice tasks.

The process may be very different, however, when subjects simply choose between options without any feedback or learning taking place at the same time. For example, Rustichini *et al.* (46) asked normal subjects to make choices among ambiguous lotteries, risky lotteries, and certain outcomes while their brain activity was monitored. Subjects were paid for the outcome of their choices, but the outcome was communicated only after the experiment was over.

Under these conditions, the VMPFC did not show any activation; it was actually less active when choices were being made than when subjects waited between trials. These results suggest that emotional circuits may be important in learning and processing information, rather than in selecting among alternatives.

Together, these data may begin to explain, in a mechanistic way, how information is analyzed when at least one class of behavior which is not predicted by the expected utility theory is produced. The process of learning and evaluating feedback may involve emotion-related areas. Ambiguity aversion, whether advantageous or disadvantageous in a particular situation, may become explicable as we learn more about the computations that brain areas like the VMPFC perform.

*Strategic cooperation.* As with the Ellsberg paradox, challenges have also been raised recently to classical game theory. In a path-breaking study, Guth *et al.* (47) analyzed the behavior of subjects playing the ultimatum game. In this game, a first player, the proposer, has \$10 to split with a second player. He can offer any amount between zero and \$10. The second player is informed of the offer and can accept or refuse. If she accepts, the split is made. If she refuses, both players get nothing. The prediction of a restrictive concept of game theory, the subgame perfect equilibrium, is that for any positive amount offered by the proposer, the second player knows that she faces a choice between gaining nothing (if she refuses the offer) or something (if she accepts). The proposer should therefore always offer the minimum possible split to player two, who should always accept. Contrary to this prediction, the robust experimental finding is that low offers (typically \$2 or even \$3) are consistently refused. The second player appears to prefer, under these conditions, to gain nothing. Anticipating this, proposers typically avoid low offers.

Although expected utility theorists have proposed some explanations for this behavior, it may well be that by analyzing the neural circuits active during the ultimatum game we may be able to both explain the causes of this behavior and to predict it. Studying the ultimatum game in subjects undergoing brain scans, Sanfey *et al.* (48) found that offers refused by the second players activated specific brain circuits in those players, and interestingly these brain circuits are also associated with emotional arousal: the anterior insula (AI, associated with disgust, both physical and emotional), the dorsolateral prefrontal

cortex (DLPFC, associated with goal maintenance and executive control), and the anterior cingulate cortex (ACC, associated with detection of cognitive conflict). Also significant is the correlation of activation with choices: An activation of the AI is positively correlated with rejection, suggesting that an emotional arousal associated with a low offer is correlated with rejection. The overall picture is that offers we might consider unfair may activate emotional circuits of the brain involved in the decision to reject an offer. If we can come to more fully understand how these circuits reach this conclusion, then a behavior that was difficult for classical game theory to predict may become fully explicable with the synthetic approach that neuroeconomics provides.

A similar line of investigation has examined interplayer cooperation during single rounds of the trust game. In this game, two players move sequentially. The first player can decide to transfer a sum of money out of an initial endowment that she receives into an investment pool that immediately triples in value. The second player then gains control of the investment and can divide it between the two players in any way he chooses. In this game, the only Nash equilibrium choice for the first player is to transfer nothing into the investment. Were she to make any transfer, the second player should take all of the money for himself. However, in real experiments the first player typically does transfer a significant amount into the investment, and the second player reciprocates by returning part of the pool. McCabe *et al.* (49) had subjects play the trust game both against a human opponent and against a computer program which, they were told, would play a human-like strategy. Under these conditions McCabe and colleagues found that subjects were more likely to cooperate with real humans than with computers and that cooperators have a significantly different brain activation in the two conditions. Cooperation is associated with activation of the anterior paracingulate cortex, a brain region associated with (50, 51) interpreting and monitoring the mental state of others.

Although these studies are in their early stages, they suggest the existence of specific brain components that make specialized contributions to decision making. The challenge that these studies face is to derive detailed computational models of the neural mechanisms, which will make neuroeconomic models broadly predictive as well as explanatory.

### Summary

Economics, psychology, and neuroscience are converging into a single, unified field

aimed at providing a theory of human behavior. In this enterprise, the method and the standard set by neuroscience is the final goal: a reconstruction of the process and mechanism that goes from a stimulus presented to the subject to his final action in response. Economics provides the conceptual structure and the object of the analysis. In this emerging view, people are seen as deciding among options on the basis of the relative desirability of each option. This is true when they are in isolation as well as when they are in strategic (interaction with few persons) and market (interaction with a large number) environments. The recent research we have been surveying describes how desirability is realized as a concrete object, a neural signal in the human and animal brain, rather than as a purely theoretical construction. Desirability is computed and is represented in the brain, and we now have the means to test, measure, and represent this activation.

But the complete reconstruction of the decision process, and hence of human behavior, is not going to be easy, because two of the cornerstones of economic analysis, subjective utility theory and Nash equilibrium, provide, even from the descriptive point of view, an incomplete picture. For example, desirability as represented by the simple economic formalism of expected utility may be appropriate only in simple conditions, where ambiguity is excluded. A more general notion is needed and, as we have seen, is beginning to be investigated and developed by psychologists and economists working together. The goal of the emerging neuroeconomic program will have to be a mechanistic, behavioral, and mathematical explanation of choice that transcends the explanations available to neuroscientists, psychologists, and economists working alone. Although it is unclear today how complete this explanation will ultimately be, neuroeconomic approaches have already begun to yield substantial fruit and to fuse natural and social scientific approaches to the study of human behavior.

### References and Notes

1. E. O. Wilson, *Consilience* (Knopf, New York, 1998).
2. D. Bernoulli, *Econometrica* **22**, 23 (1954).
3. D. Kahneman, P. Slovic, A. Tversky, Eds, *Judgement Under Uncertainty: Heuristics and Biases* (Cambridge Univ. Press, Cambridge, 1982).
4. C. S. Sherrington, *The Integrative Action of the Nervous System* (Scribner, New York, 1906).
5. I. P. Pavlov, *Conditioned Reflexes an Investigation of the Physiological Activity of the Cerebral Cortex* (Oxford Univ. Press, Oxford, 1927).
6. T. Caraco, S. Martindale, T. Whittam, *Anim. Behav.* **28**, 820 (1980).
7. A. Kacelnik, M. Bateson, *Trends Cogn. Sci.* **1**, 304 (1997).
8. A. Robson, *J. Polit. Econ.* **109**, 900 (2001).
9. H. B. Barlow, in *Current Problems in Animal Behavior*, W. H. Thorpe, O. L. Zangwill, Eds. (Cambridge Univ. Press, Cambridge, 1961).
10. P. W. Glimcher, *Decisions, Uncertainty and the Brain: The Science of Neuroeconomics* (MIT Press, Cambridge, MA, 2003).
11. C. S. Sherrington, *The Integrative Action of the Nervous System* (Scribner, New York, 1906).
12. W. T. Newsome, K. H. Britten, J. A. Movshon, *Nature* **341**, 52 (1989).
13. W. T. Newsome, K. H. Britten, C. D. Salzman, J. A. Movshon, *Cold Spring Harbor Symp. Quant. Biol.* **55**, 697 (1990).
14. See also (50) and (51) for a parallel set of studies that presaged and continues to parallel many of these developments.
15. M. N. Shadlen, K. H. Britten, W. T. Newsome, J. A. Movshon, *J. Neurosci.* **16**, 1486 (1996).
16. R. A. Andersen, L. H. Snyder, D. C. Bradley, J. Xing, *Annu. Rev. Neurosci.* **20**, 303 (1997).
17. C. L. Colby, M. E. Goldberg, *Annu. Rev. Neurosci.* **22**, 319 (1999).
18. M. L. Platt, P. W. Glimcher, *Nature* **400**, 233 (1999).
19. B. Knutson, C. M. Adams, G. W. Fong, D. Hommer, *J. Neurosci.* **21**, RC159 (2001).
20. M. P. Paulus, N. Hozack, X. Zauscher, *Neuroimage* **13**, 91 (2001).
21. H. C. Breiter, I. Aharon, D. Kahneman, A. Dale, P. Shizgal, *Neuron* **30**, 619 (2001).
22. D. Kahneman, A. Tversky, *Econometrica* **47**, 263 (1979).
23. J. I. Gold, M. N. Shadlen, *Nature* **404**, 390 (2000).
24. J. I. Gold, M. N. Shadlen, *Trends Cogn. Sci.* **5**, 10 (2001).
25. M. N. Shadlen, K. H. Britten, W. T. Newsome, J. A. Movshon, *J. Neurosci.* **16**, 1486 (1996).
26. M. N. Shadlen, W. T. Newsome, *J. Neurophysiol.* **86**, 1916 (2001).
27. L. P. Sugrue, G. S. Corrado, W. T. Newsome, *Science* **304**, 1782 (2004).
28. P. Janssen, M. N. Shadlen, *Soc. Neurosci. Abstr.* **767**, 2 (2003).
29. B. Coe, K. Tomihara, M. Matsuzawa, O. Hikosaka, *J. Neurosci.* **22**, 5081 (2002).
30. J. V. Von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton Univ. Press, Princeton, NJ, 1944).
31. J. F. Nash, *Proc. Natl. Acad. Sci. U.S.A.* **36**, 48 (1950).
32. J. F. Nash, *Econometrica* **18**, 155 (1950).
33. D. J. Barraclough, M. L. Conroy, D. Lee, *Nat. Neurosci.* **7**, 404 (2004).
34. M. C. Dorris, P. W. Glimcher, *Neuron*, (in press).
35. J. F. Nash, *Ann. Math.* **54**, 286 (1951).
36. P. R. Montague *et al.*, *Neuroimage* **16**, 1159 (2002).
37. D. Ellsberg, *Q. J. Econ.* **75**, 643 (1961).
38. M. Allais, *Econometrica* **21**, 503 (1953).
39. For a review of these issues, see D. Luce, *Utility of Gains and Losses*, vol. 8 of *Scientific Psychology Series* (L. Erlbaum, Mahwah, NJ, 2000).
40. G. Gigerenzer, R. Selten, *Bounded Rationality: The Adaptive Toolbox* (MIT Press, Cambridge, MA, 2002).
41. J. Dickhaut *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3536 (2003).
42. See for example A. Bechara, H. Damasio, D. Tranel, A. Damasio, *Science* **275**, 1293 (1997).
43. J. O'Doherty, M. L. Kringelbach, E. T. Rolls, J. Hornak, C. Andrews, *Nat. Neurosci.* **4**, 95 (2001).
44. A. Rustichini, J. Dickhaut, P. Ghirardato, K. Smith, J. Pardo, in *Games and Economic Behavior*, in press.
45. W. Guth, R. Schmittberger, B. Schwarze, *J. Econ. Behav. Organ.* **3**, 347 (1982).
46. A. Sanfey, J. K. Rilling, J. Aronson, L. E. Nystrom, J. Cohen, *Science* **300**, 1755 (2003).
47. K. McCabe, D. Houser, L. Ryan, V. Smith, T. Trouard, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11832 (2001).
48. C. D. Frith, U. Frith, *Science* **286**, 1692 (1999).
49. H. L. Gallagher, C. D. Frith, *Trends Cogn. Sci.* **7**, 77 (2003).
50. C. R. Gallistel, *Cognition* **50**, 151 (1994).
51. K. Conover, P. Shizgal, in *Games and Economic Behavior*, in press.