Strategizing in the Brain

Colin F. Camerer

Most economic theories minimize the influence of human emotions and assume that what people believe and choose follows rationality principles. Important principles include knowing how much of one valuable good is worth one unit of another; following the rules of probability in processing information; planning ahead; resisting temptation; and guessing accurately what others will do. These principles have proved useful as mathematical building blocks for devising aggregate theories of corporate and market behavior. An emerging field of study called "behavioral economics" takes advantage of dramatic advances in psychology and neuroscience. Behavioral economics replaces strong rationality assumptions with more realistic ones and explores their implications (*1*). A clear demonstration of how neural evidence contributes to behavioral economics is provided by Sanfey *et al.* on page XXXX of this issue (*2*). These investigators analyzed subjects with functional magnetic resonance imaging as they played the "ultimatum game", and correlated activity in certain brain areas with the cognitive and emotional processes involved in economic decision-making.

Enhanced online at www.sciencemag.org/cgi/content/full/300/5627/1673

The author is in the Division of Humanities and Social Sciences 228-77, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: camerer@hss.caltech.edu

In their version of the ultimatum (or "take-it-or-leave-it") game, a proposer offers a division of $10 to a responder, who accepts it (ending the game) or rejects it, leaving both players with nothing. The prediction is that responders who want to earn the most money will take any offer; a self-interested proposer who anticipates this will offer the smallest amount. However, this prediction turns out to be wrong (offers are typically around 50% of the total amount, and 50% of low offers are rejected). Functional magnetic resonance imaging shows why. Subjects whose brains were imaged while they were presented with an unfair offer ($1 to $2 out of the $10 available) showed greater activity in the bilateral anterior insula of the brain revealing that such an offer created negative emotions. Those subjects with the strongest activation of the anterior insula rejected a higher proportion of unfair offers. The anterior cingulate (ACC), a brain region that detects cognitive conflict, also showed greater activity during presentation of an unfair offer suggesting that this area mediates the conflict between earning money and feeling bad. These findings emphasize the importance of emotional influences in human decision-making.

In games like ultimatum bargaining, "players" with information choose "strategies" that, collectively, create outcomes which players like or dislike and to which they attach numerical valuations ("utilities"). Game theory can link economics to other sciences because it uses the same tools to model interactions at many different scientific levels (genes, firms, nation-states). But doing so requires extending the central assumptions of rational game theory, namely, that players are (i) self-interested and (ii) reach an "equilibrium" in which everyone is choosing (or planning) strategies that yield the best outcome, anticipating that others are doing the same. An emerging approach called "behavioral game theory" replaces these assumptions with precise alternatives that are more cognitively plausible (*3*).

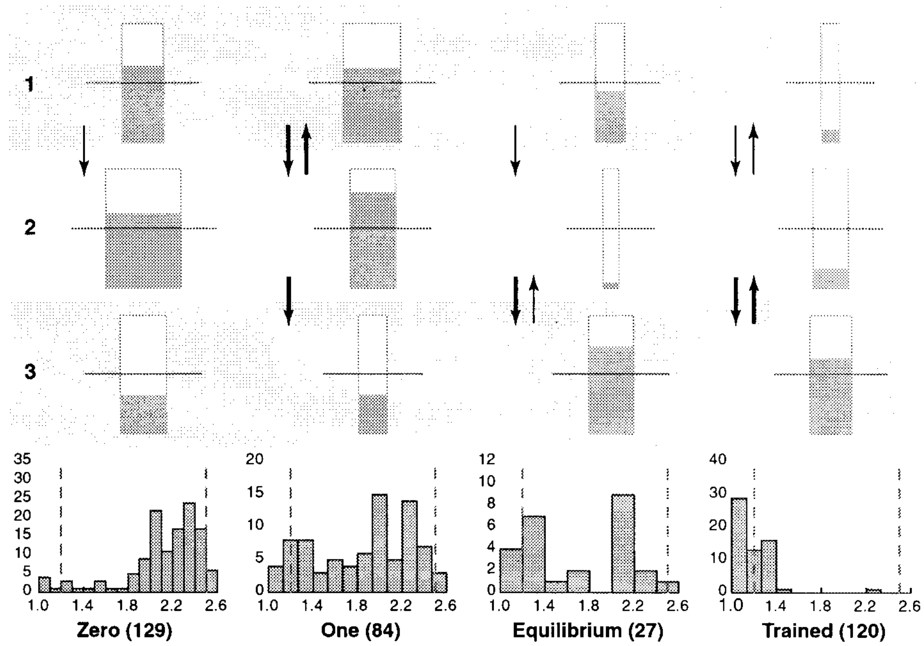
One ingredient of behavioral game theory is a model of "social utility," showing how players' utilities for payoff allocations depend on how much others get, as well as on their own payoffs. This old idea is illustrated by many experiments showing that people routinely cooperate in the "prisoner's dilemma" game, in which "defection" is always better for one player but mutual "cooperation" makes everyone in a group better off. Prisoner's dilemma cooperation is well established, but evidence from newer games (like ultimatum bargaining) help to calibrate precise social utility functions that make fresh predictions (*4*).

One new game is "trust." In a trust game, one player can invest money that is multiplied (say, by 3) to reflect the investment's productivity. A second trustee player can repay or keep as much of the tripled investment as he or she likes. The baseline game is played with anonymous strangers, as a model of investment in economies with "country risk" and poor legal protection for investors, or firms paying wages and trusting employees to work hard (4). When the trust game is played once, the first player usually invests about half and is repaid roughly what he or she invested (3). Trustworthiness correlates with the trustee's level of oxytocin after being trusted, a hormone important in social bonding (5), and also correlates with economic growth across economies. Mutual cooperation in the prisoner's dilemma also leads to increased cortical activation, suggesting a neural location for the "warm glow" of receiving reciprocation (6). When the trust game is played a fixed number of times, players invest and repay in early periods, but trust breaks down when the repeated game's end grows near (3).

Theories of social utilities that include guilt from getting more than others, and envy from getting less, predict rejections of low ultimatum offers to reduce envy, and repayment of trust to reduce guilt (3, 4). The Sanfey *et al.* results provide a neural foundation for these models, by locating distaste for envy in the insula, and the trade-off between envy and money in the ACC. Competing theories focus on the human instinct to reciprocate, and the idea that one-shot games are an "unnatural habitat" for humans who evolved playing repeated games in close-knit groups (3).

The rejections Sanfey *et al.* report in their ultimatum game are typical. But results from special populations illuminate the boundary of this regularity and its causes. Autistic adults (7) and players in some small-scale societies (8) offer less and accept very little, which suggests, ironically, that rational game theory only works in ultimatum games when deficits in "mind reading" or cultural adaptation lead to extremely uneven offers. Cooperativeness in normal subjects is also correlated with heightened activity in the prefrontal cortex (Brodmann area 10), implicated in mind reading, and with limbic activity (9).

A second ingredient of behavioral game theory is the idea that players use only one or two steps of iterated strategic reasoning (3, 10). Limited strategizing is illustrated by the "beauty contest" game, in which players choose numbers from 0 to 100, and the player whose number is closest to two-thirds of the average wins a fixed prize. Rational game theory predicts an equilibrium in which choices are mutually consistent (leading to 0), but in experiments many players assume that others are making random choices, and choose 33, or believe that others are responding to random choices, and choose 20 to 35. Models of this cognitive hierarchy organize regularity from many games and show that one or two steps of thinking are typical, although three or four steps are used by analytically skilled undergraduates and game theorists (10, 11).



Planning ahead when bargaining. The diagram explains the workings of a "bargaining game" equivalent to a stretched-out version of the ultimatum game described by Sanfey *et al.* (2). The four columns of vertically arrayed boxes (white) each represent the payoff in one stage of a three-stage bargaining game with alternating offers (top box is round 1, bottom box is round 3). In stage 1 (top box) players divide \$5; in stage 2 they divide \$2.50; in stage 3 they divide \$1.25 (if the third-stage offer is rejected they earn nothing). The shaded regions of boxes show the relative time in which each box is clicked open to reveal the payoff in that stage, the relative number of times each box is opened (box width), and the relative number of forward (down arrow) and backward (up arrow) transitions between stages (arrow thickness; average transitions of less than one per trial omitted). Columns (from left) show what happens when players look most often at the first box (level 0 thinking), look more often at the second box than the first (level 1 thinking), look more often at the third and second than the first (equilibrium), and when subjects are trained in backward induction (trained). Bottom histograms show frequency distributions of offers from players whose looking times correspond to each of the four patterns. Dotted vertical lines show an even \$2.50 offer (right vertical line) and the game-theoretic ("perfect") equilibrium \$1.25 offer (left vertical line). Looking more often at later-stage boxes (in columns 2 and 3) shifts offer distributions to the left, toward the rational game theory prediction of \$1.25.

A third ingredient of behavioral game theory is learning: Simple parametric models of generalized reinforcement explain learning paths in dozens of games (12). These models are supported by studies showing that accumulated reinforcement correlates with single-neuron parietal activity in monkeys making rewarding eye saccades (13). But some human players are "sophisticated"; they realize that others are learning and choose strategies that build beneficial reputations, like repaying early in trust games that are repeated (3, 11).

Studies of alternating-offer bargaining illustrate all three ingredients. Consider a three-stage game in which the players first bargain over a \$5 "pie." If the first player's offer is rejected, the second player gets to make a counteroffer dividing \$2.50; if that offer is rejected player 1 makes a final (ultimatum) counteroffer dividing \$1.25 (see the figure) (14). Game theory predicts that the self-interested first player who plans ahead will offer \$1.25. In experiments, opening offers are more generous, \$2 to \$2.50, and lower offers are rejected half the time. Measuring attention in computer displays shows that most players barely look ahead, or look only one stage ahead. Players who spontaneously look further ahead or are trained to do so make offers closer to the rational game theory prediction.

Constrained by these facts, game theory is trifurcating into three approaches: rational game theory, useful for modeling firms and countries that pool cognitive resources and hire consultants to analyze games carefully (for example, bidding for telecommunications spectrum in auctions); behavioral game theory, useful for explaining what normal people do and how they learn, which is important for explaining strikes, divorces, incentive contracts, and litigation; and evolutionary game theory, which explains equilibration in animal populations by natural selection, and imitation among humans (social selection).

Informing game theory is only one way in which neuroscience will reshape economics. Other puzzles that neuroeconomics can help to solve include the fact that people find some trade-offs between health, time, statistical lives, clear air, and money difficult or offensive; how advertising works; the fact that many people do not plan lifetime savings optimally, or succumb to ruinous temptation and addiction; and the fact that financial markets and macroeconomies fluctuate so dramatically (*15*).

References

1. C. F. Camerer, G. Loewenstein, in *Advances in Behavioral Economics*, C. Camerer, G. Loewenstein, M. Rabin Eds. (Princeton Univ. Press, Princeton, NJ, 2003).
2. A. Sanfey *et al*, *Science* **300**, XXX (2003).
3. C. F. Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton Univ. Press, Princeton, NJ, 2003).
4. E. Fehr, S. Gächter, *J. Econ. Persp.* **14**, 159 (2000).
5. P. Zak, R. Kurzban, W. T. Matzner, in preparation.
6. J. K. Rilling *et al.*, *Neuron* **35**, 395 (2002).
7. E. Hill, D. Sally, *Journal?*, in preparation.
8. J. Henrich *et al.*, *Am. Econ. Rev.* **90**, 973 (2002).
9. K. McCabe *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11832 (2001).
10. M. Costa-Gomes *et al.*, *Econometrica* **69**, 1193 (2001).
11. C. F. Camerer *et al.*, *Am. Econ. Rev.*, in press.
12. C. F. Camerer, T. Ho, *Econometrica* **67**, 827 (1999).
13. P. Glimcher, *Decisions, Uncertainty, and the Brain: The New Science of Neuroeconomics* (MIT Press, Cambridge, MA, 2003).
14. E. Johnson *et al.*, *J. Econ. Theory* **104**, 16 (2002).