Neuron
NeuroView

Neuroeconomics: Opening the Gray Box

Colin F. Camerer[1,*]
[1]Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA
*Correspondence: camerer@hss.caltech.edu
DOI 10.1016/j.neuron.2008.10.027

The long-run goal of neuroeconomics is to create a theory of economic choice and exchange that is neurally detailed, mathematically accurate, and behaviorally relevant. This theory will result from collaborations between neuroscientists and economists and will benefit from input from other fields, including computer science and psychology.

Neuroeconomics is a combination of mathematical frameworks, experimental paradigms, and lab and field behavioral data about peoples' choices (from economics) and measures of neural activity (from neuroscience). The goal is to relate mathematical theories of choice to neural measures, to build hypotheses that constrain competing economic theories, to predict effects of cognitive and emotional factors on individual choices, and to suggest when people do not always choose what is best for them (and what good policies follow).

How Economic Models Work

Imagine entering a bookstore looking for vacation reading. There are an enormous number of choices (a typical Border's might have 100,000). The cover art, size, font, and heft of books all invoke sensory processing that might influence what you buy. Picking up a possible purchase, memories about similar books that were loved and hated are called up by an internal self-recommendation system. Was it recommended by Oprah or by friends? (Social processing kicks in.) If it's an unknown author, there is a vague risk that the book is awful. If it is a hardback by a favorite author, is it worth waiting a while for the cheaper paperback to save some money?

Theories in economics generally start with the presumption that choices like these, from different sets of books, are consistent in special mathematical ways. For example, if you're holding book A you should not then trade A for B, B for C, and trade C back for A again (exhibiting an "intransitive cycle"). If you can always make up your mind between two books, and don't choose in a cycle, then you are acting as if you are implicitly assigning ordered numbers to all the books and picking the higher-numbered book.

Economists call these implicit numbers "utilities." The theory of utility-maximization says people make the best choices given what they want, know, and can afford. This "preferences-information-constraints" mathematical framework can be applied to a wide range of market and political choices, including social ones in which the choices of other people influence a person's behavior (game theory). While utility-maximization might appear to be an implausible interpretation of what people shopping in the bookstore are literally thinking and doing, it makes predictions that are correct in a wide range of settings (e.g., when the price of a good goes up people buy less of it), even for many animal species.

Economists are proudly agnostic about the neural basis of utilities: if we are unable to find measures of neural activity that correlate with numerical utilities, most economists would not abandon the theory. Given the agnostic stance, what can neuroscience add (Bernheim, 2008)? Some studies will show neural circuitry that *does* appear to implement utility computation. Knowing how that circuitry works will have some novel implications. Such knowledge could also establish the limits of consumers' cognition in understanding the consequences of their choices, and inform policy.

Neuroscience probably has more to contribute to understanding consumer decision making (demand) than to understanding the supply of goods, except for some topics like how emotions, norms, and rewards motivate workers, how job skills ("human capital") develop in the brain, and how service experiences are valued.

Toward a General Framework

Figure 1 contrasts neuroeconomic elements of choice with constructs in "behavioral economics" and standard economics. Behavioral economics uses psychological facts and constructs to incorporate limits on computational ability, willpower, and self-interest in economic analysis. These limits are often incorporated by adding a behavioral parameter or process to a standard economic model. A close correspondence between neural activity and *some* version of economic theory (including extra behavioral processes) is already emerging—tentatively, of course. The diagram suggests at least three fundamental questions.

(1) In neuroscientific models, stable behavioral choice patterns are the *end result* of learning; in economic models, they are the *starting point* of analysis. This contrast raises a question: when are learning processes consistent with choice patterns reflective of stable preferences?

One answer is that rational economic theories probably are good approximations when choices are simple and repeated in a stationary environment, so that goal and decision values can be learned. If you eat regularly at a local restaurant for many years and sample most of the dishes, an unchanging menu may become like a list of conditioned stimuli that evoke stable valuation signals. Your menu choices might then satisfy economic consistency axioms to a surprising degree. In such choices, a form of learning referred to as temporal difference (TD) learning is often thought to apply. In TD learning, the gap between the actual outcome and a person's expectation of value, called the "prediction error," is used to adjust the estimate of goal value. Mathematical analyses show that TD

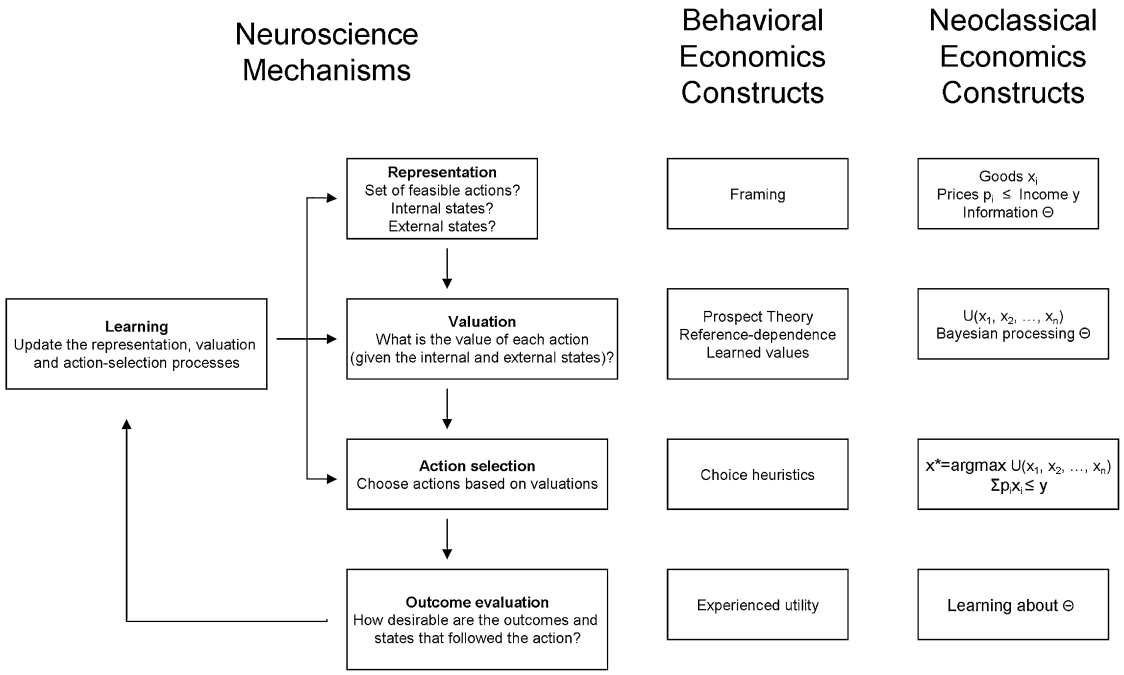

Figure 1. Neural and Economic Components of Choice
Contrasting neuroscientific mechanisms of choice (left) with additional behavioral economics constructs (middle) and basic neoclassical economics constructs (right). Adapted from *Nature Reviews Neuroscience* (Rangel et al., 2008).

learning can be used to learn even complex goal values and strategies (e.g., a leading backgammon algorithm uses TD learning).

Many studies have shown neural correlates of TD learning during goal-oriented choice. Early evidence showed that neural firing rates in midbrain dopaminergic neurons, which project to the striatum and prefrontal cortex (Schultz et al., 1997), appear to encode TD error. fMRI signals of TD errors were later found in the striatum (O'Doherty et al., 2003). The striatal regions are active when people imagine possible counterfactual outcomes ("fictive learning," Lohrenz et al., 2007) and when rewards are social (Fehr and Camerer, 2007).

TD learning will generate reliable stable goal values, which sets the stage for showing a strong "neurometric" correlation between measures of neural activity and the economists' inferred utility value of such goods in practiced or familiar choices. Neural firing rates from single-unit recordings correlate with the inferred value of choices in risky games (Platt and Glimcher, 1999) and with behavioral choices between different juice rewards (Padoa-Schioppa and Assad, 2006). Other studies show very reliable correlations between the strength of fMRI signals in human medial OFC and values assigned to foods, wine, charitable giving, and consumer goods (Hare et al., 2008). Despite the assertion of agnostic economists that people only act "as if" they maximize utility, these studies suggest that there really are neural measures that deserve to be called utilities!

However, many complex decisions are so infrequent that convergence is unlikely to occur through TD learning. Shopping for books, for example, involves integration of sensory, abstract, and memory processing, social influence, budget constraint, and tradeoffs about risk (new authors) and time (waiting for paperbacks). And new books are always coming out. Choosing a house or college is even more complex and uses abstract propositional inputs. A challenge for neuroeconomics is to show what neural systems collaborate in these types of complex, high-stakes, slowly learned decisions.

(2) How do internal and external constraints influence choice? The "representation" box in Figure 1 includes both external and internal constraints. In economic theory, the typical external constraints are prices of goods (interpreted broadly to include time and social factors) and a consumer's income. However, in standard economic theory, it is rare that *any* internal constraints are made explicit. Consumers are usually assumed to choose, given their preferences and information, optimally and unemotionally. However, economists have a concept of "state dependence" of utility—e.g., the demand for food goes up when you are in a hunger state. However, the states are usually assumed to affect preferences and are not under cognitive control.

Cognitive neuroscience could illuminate the influence of internal constraints on economic choice. An example is emotional regulation. Subjects instructed to "reappraise" fearful stimuli in a way that allows them to consciously dampen their fear responses show less activity in amygdala and more activity in the prefrontal cortex (two parts of an apparent regulatory circuit). This cognitive reappraisal paradigm has been extended to economic choices of risky gambles; preliminary evidence suggests that downregulation of loss aversion reduces the tendency to avoid gambles that may yield losses and reduces skin conductance (suggesting

a reduction in negative emotional response). There are many more researchable examples from bargaining, bluffing, and selling. In these cases, a cognitive process can be both "chosen" and then influence valuation of choices.

(3) One of the most solid findings in the neuroeconomics of reward is that dissociable systems guide three distinctly different types of valuation: Pavlovian conditioning systems (learning to associate a particular conditioning cue with later reward), habit systems, and goal-directed systems. A basic question is how do these different valuation systems work and interact?

Standard economic theory does not explicitly recognize these different types of valuation (though the theory could be extended to include them). Their differences can have important economic consequences. For example, the habit system is likely to be slow to respond to abrupt changes, such as a sharp jump in the price of gas. For the habit system, a sudden change in an item's price requires a change in our behavioral response that is analogous to "reversal learning." Economists do typically distinguish short-run and long-run responses to large changes, but this distinction is rarely linked to details of learning and therefore has little predictive power.

Conversely, the goal-directed valuation system will typically involve a lot of top-down processing of numerical and abstract concepts (or should!), such as "crunching the numbers" in valuing complex decisions (e.g., buying solar panels). This type of processing implies that consistency of choices is likely to be modulated by variables related to expertise, cognition, attention, and so forth.

What happens when valuation systems conflict? Economists have approached this question using "dual-system" models, e.g., one model of addiction (Bernheim and Rangel, 2004) distinguishes "hot" states, in which habit rules, from "cold" states, in which choice is goal directed. In the cold state, people can either deliberately invest in kicking an addiction or can take a chance on whether a hot state will develop and result in craving. Since these dual-system models sometimes create analogies between two types of interacting economic agents and two similar types of brain processes, neural tests of them provide the most direct evidence.

Neural Evidence for Behavioral Economics

Some progress has already been made in finding neural correlates of behavioral economics models of time preference, social reward, strategic thinking, and risky choice. For brevity, I will only discuss the latter.

In standard economic models, gambles have possible outcomes x with associated probabilities p(x). Gambles could be monetary ones, like lottery tickets, or de facto lotteries like going to college hoping to earn more, starting a small business, or running a quick errand without feeding a parking meter.

A normatively appealing theory, called "expected utility," is that such gambles are valued by weighting the hedonic utility of possible outcomes by the chances of those outcomes actually occurring. A modified view, called "prospect theory," differs in two ways: probabilities are not always weighted by their numerical value, and outcomes are valued relative to a reference point. There is much lab evidence and some field evidence (e.g., from game shows and stock prices) consistent with prospect theory and emerging neuroscientific evidence.

Choices do appear to implicitly weight probabilities nonlinearly. Preliminary evidence from Hsu et al. using fMRI shows striatal valuation signals that imply that a one-in-a-million chance has a neural weight of 0.02. This process is consistent with economic overreactions to tiny chances of blissful outcomes (e.g., winning a lottery) and rare catastrophes (e.g., plane crashes).

Prospect theory also proposes that choice outcomes have a "reference-dependent" value $v(x - r)$ relative to a point of reference along with an absolute utility $u(x)$. The value $v(x - r)$ is thought to reflect hedonic adaptation to the past, or valuation relative to a future aspiration. However, this reference-dependent component of "value" might be somehow related to a learning signal (a la TD learning).

Reference dependence implies that choices may reverse when a natural point of reference is reversed (a "framing" effect). De Martino et al. (2006, 2008) find that lateral OFC activity is correlated with reversal tendencies across subjects, and reversals are also less common in autistic subjects. Another implication is that "losses" relative to a reference point ($x < r$) and "gains" ($x > r$) could have different neural bases and implications. Indeed, many studies with humans (and some with monkeys) find that the pattern of observed choices implies an aversion to loss—losses are valued about twice as large as equal-sized gains. This "loss aversion" correlates with differential brain activations to increased gain and reduced loss using fMRI (Tom et al., 2007). Still another implication is that owning goods creates a reference point, so that selling goods is a distasteful loss, which creates an "endowment effect" compared to buying unowned goods. fMRI evidence suggests that this effect is not due to an exaggerated valuation, but to a heightened sensation of unpleasant loss when selling (Knutson et al., 2008).

These studies show how prospect theory parameters, inferred mathematically from behavior, often correlate with neural activity in regions generally thought to be involved in valuation judgments or learning. These neurometric correlations can be used to make behavioral economic predictions that can be validated by lesion patient studies, TMS, and single-unit recording and stimulation. Knowing how a brain region's activity is linked to behavioral parameters also invites predictions about how decisions would change across the lifecycle, across genetic backgrounds, and in response to pharmacological intervention.

Conclusion

Economists have mixed feelings about neuroeconomics. Many think direct neural evidence is unnecessary. Others share the skepticism of cognitive psychologists (and many neuroscientists) about how rapidly techniques like fMRI will yield surprising conclusions about complex economic choice processes. Still others are cautiously optimistic and think that exploring new technologies has option value.

My view is that the largest long-run impact will come from ways in which neuroscience challenges the preference-information-constraint framework by showing the influence of internal constraints and cognitive variables.

For example, part of economic theory is the idea that inferences are often logical and consistent. But neural systems evolved (in a series of kludges) to solve adaptive

problems. Neuroscience will show more clearly conflicts in which behavior is biologically plausible rather than logical.

To illustrate, in standard theories people should not care whether a probabilistic risk is well understood or not and should not care whether outcomes are framed as gains or losses from a reference point (holding their final consequences fixed). Ironically, those two patterns are exhibited most strongly by lesion patients with lateral OFC damage (Hsu et al., 2005) and by autists (De Martino et al., 2008)—that is, the most "logical" behavior is exhibited by abnormal people. These results suggest a shift from logical criteria for economic rationality to biological ones.

Economic analyses take preferences as a starting point and have little to say about changes in preference. A deeper understanding of genetics, learning, and childhood development will provide some ideas about how preferences are related (e.g., are patient people less averse to risk?) and how preferences for food, violence, saving, risk taking, and education vary across people and change with experience and development.

What can economics do for neuroscience? Economic theories supply a parametric language for linking choices to components of valuation. Even better, those ideas have been extended to social exchange, in the form of game theory. And as noted earlier, economists now have several models in which two or more types of agents interact, which can be starting points for improved models of multiple systems in the brain.

REFERENCES

Bernheim, B.D. (2008). In Neuroeconomics: Decision Making and the Brain, P. Glimcher, et al., ed. (New York: Elsevier), pp. 115–126.

Bernheim, B.D., and Rangel, A. (2004). Am. Econ. Rev. *94*, 1558–1590.

De Martino, B., Kumaran, D., Seymour, B., and Dolan, R. (2006). Science *313*, 684–687.

De Martino, B., Harrison, N., Knafo, S., Bird, G., and Dolan, R. (2008). J. Neurosci. *28*, 10746–10750.

Fehr, E., and Camerer, C.F. (2007). Trends Cogn. Sci. *11*, 419–427.

Hare, T.A., O'Doherty, J., Camerer, C.F., Schultz, W., and Rangel, A. (2008). J. Neurol. *28*, 5623–5630.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., and Camerer, C.F. (2005). Science *310*, 1680–1683.

Knutson, B., Wimmer, G.E., Rick, S., Hallon, N.G., Prelec, D., and Loewenstein, G. (2008). Neuron *58*, 814–822.

Lohrenz, T., McCabe, K., Camerer, C.F., and Montague, P.R. (2007). Proc. Natl. Acad. Sci. USA *104*, 9493–9498.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., and Dolan, R.J. (2003). Neuron *38*, 329–337.

Padoa-Schioppa, C., and Assad, J.A. (2006). Nature *441*, 223–226.

Platt, M.L., and Glimcher, P.W. (1999). Nature *400*, 233–238.

Rangel, A., Camerer, C., and Montague, P.R. (2008). Nat. Rev. Neurosci. *9*, 545–556.

Schultz, W., Dayan, P., and Montague, P.R. (1997). Science *275*, 1593–1599.

Tom, S., Fox, C., Trepel, C., and Poldrack, R. (2007). Science *315*, 515–518.