Economics and Philosophy, **24** (2008) 369–379
doi:10.1017/S0266267108002022

THE POTENTIAL OF NEUROECONOMICS

Colin F. Camerer
California Institute of Technology

The goal of neuroeconomics is a mathematical theory of how the brain implements decisions, that is tied to behaviour. This theory is likely to show some decisions for which rational-choice theory is a good approximation (particularly for evolutionarily sculpted or highly learned choices), to provide a deeper level of distinction among competing behavioural alternatives, and to provide empirical inspiration for economics to incorporate more nuanced ideas about endogeneity of preferences, individual difference, emotions, endogeneous regulation of states, and so forth. I also address some concerns about rhetoric and practical epistemology. Neuroscience articles are necessarily speculative and the science has proceeded rapidly because of that rhetorical convention. Single-study papers are encouraged and are necessarily limited in what can be inferred, so the sturdiest cumulation of results, and the best guide forward, comes in review journals which compile results and suggest themes. The potential of neuroeconomics is in combining the clearest experimental paradigms and statistical methods in economics, with the unprecedented capacity to measure a range of neural and cognitive activity that economists like Edgeworth, Fisher and Ramsey daydreamed about but did not have.

Debating the *potential* of neuroeconomics is much less interesting than debating the results and interpretations of particular studies. Furthermore, there is actually little disagreement that there is potential in understanding how the brain makes economic choices.

My comments will simply reiterate some of the ideas about what neuroeconomics is trying to do, mention some possible results, and then directly address some specific assertions which are misguided.

The goal of neuroeconomics is to build up a mechanistic, behavioural and mathematical theory of choice and exchange. To clarify, "behavioural" means observed choices. "Mechanistic" means at some level of neural circuitry, which includes psychophysiological measurement (skin conductance, heart rate, facial muscle movement, startle eyeblink responses, etc.), tracking of eye movements, and any other biological measure. Response times are surely useful too, and the easiest to measure though they are actually a choice of time rather than a choice; I am certainly glad that Rubinstein (2008, this issue) is one of the many people gathering these data and putting them to work.

Note that economics (in the form of revealed preferences and the myriad axiomatic foundations of decision) already *has* a mathematical and behavioural theory, but the theory is usually interpreted as agnostic about the mechanistic details.[1] Conversely, many neuroscience investigations compare behaviour and speculate about mechanism but do not use the language of computation or mathematics to link the two.

For most economists, the ideal contribution of neuroeconomics would go further than showing a mechanistic basis for behaviour expressed in mathematical form. The biggest challenge is to show a neural basis for some kinds of choices or judgments, which predicts how an observable variable would influence choices in a reduced-form model in a surprising new way.

For example, suppose one found that activity in a brain region R was correlated with how quickly information is reflected in prices in an experimental market. (Ideally, one would establish even more by showing that markets with subjects who are patients with damage to R have slower information aggregation, or that TMS disruption of region R slows information aggregation, or that adding a distractor task which activates R changes the rate of information aggregation.) Using what is known about the types of tasks or stimuli that activate R (that's one of the hard parts), one can then hypothesize that an observable event or condition in the world would activate R (that's the other hard part) and would slow down information aggregation in actual markets. Some neuroeconomists are thinking about this but have not produced a knockout example yet.

[1] It is useful to point out, by the way, that the foundation of choice was not always considered only a matter of linking unobserved constructs to observed choice and ignoring mechanism. Ramsey, Fisher and Edgeworth were among the classical economists who fantasized about measuring hedonic bases of utility directly (Ramsey spoke of a "psychogalvanometer" and Edgeworth of a "hedonimeter"; see Colander, 2008). Admittedly, they were mostly interested in being able to establish a cardinal measure of utility, which is far from the focus of modern neuroeconomics. Given how interested these early economists were in direct measurement – when they could not readily do so – one might think that their intellectual descendants would be at least as interested when tools like those they fantasized about are actually available.

What could neural mechanisms of measurement tell us about a good mathematical foundation of observed behaviour?

I see three possibilities: The mechanisms could appear to implement rational-choice theories (perhaps for some types of decisions or for some people, such as experts); the mechanisms could appear to favour some kinds of behavioural economics alternatives over others; or the mechanisms could inspire us to think about variables that are likely to be important but do not currently have a central place in standard economic theory. These hidden variables might include willpower, attention, emotion, indecision, effects of advertising, etc., as in Spiegler's thoughtful comment (2008, this issue) and Harrison's interesting section 4 (2008, this issue).

Mechanisms of rational choice: One possibility is that some kinds of choices will turn out to be mechanistically implemented in ways that match familiar math and observed behaviour (e.g. Bayesian models of integration of sensory and motor activity are popular now among neuroscientists).

Mechanisms of behavioural alternatives: Another possibility is that understanding the mechanistic details might help inform debate about which mathematical models of behaviour are correct. As many readers will know, there is a lively debate within the economics profession about which mathematical models of which observed behaviour are on the right track, in the form of "behavioural economics" alternatives to normative rational-choice principles.

For example, expected utility theory is one mathematical form of combining risks and outcomes; much of its empirical appeal comes from the normative appeal of the independence axiom. Another is prospect theory, which takes its shape from psychophysical principles (sensitivity to a point of reference, and diminishing marginal sensitivity of value and probability weight when moving away from local reference points). There is still a very lively debate about which theory gives the best account of which types of data. Prospect theory is the most widely cited empirical article published in economics from 1970–2005 (Kim, Zingales and Morse 2007) but few graduate preliminary exams in decision theory include questions about its details. The available neural evidence thus far is supportive of the hypothesis that brain activity responds to framing of gambles, loss and gain differentials, and nonlinear weighting of probability (see Fox and Poldrack in press).

Another example is the study of ambiguity-aversion – the reluctance to act in the face of missing information or uncertainty about probability. This has been modelled in many, many different ways, including multiple priors, set-valued priors, hierarchies of second-order belief, and mean-dispersion preferences (e.g. Grant and Polak 2007). I believe it is safe to say that empirically separating these theories with naturally occurring data would be very difficult, although trying to do it is an excellent research topic.[2]

However, *if* the theories were described not only in terms of their axioms and the functional representations of choice implied by those axioms (the math, and the behaviour), but *also* in terms of their brain implementation, then one could test among the theories using brain data. Note that the point of such tests is not to establish the neuroscience of ambiguity-aversion per se (although that may interest neuroscientists). The point is to use brain evidence to adjudicate empirically among theories which are particularly difficult to distinguish using the market-prediction test (for ordinary types of data). The brain data is a way to keep pace with the remarkable ability of economic theorists to invent new mathematical foundations for a small set of observed phenomena.

My paper with Hsu *et al*. (2005) had this intent. In ambiguity and risk choices, we find some evidence for activity in response to "mean" expected utility in the striatum, and a strong response to ambiguity (compared to risk) in lateral OFC, dorsolateral PFC, and the amygdala. The fact that different regions are activated differently by the mean and by the degree of ambiguity suggests that a model in which these two elements are evaluated separately is on the right track.

However, this interpretation does not get far unless we have some idea from dozens of other studies – or in some areas of neuroscience, hundreds – about what general functions these areas seem to perform.

The striatum is reliably activated by many types of reward (money, cocaine, attractive pictures, curiosity-provoking trivia questions) so it is plausible that the striatum would also activate a sense of mean reward (or more likely, prediction error). The amygdala is thought of as a "vigilance" area that responds to fear and potential threat (though it is sometimes activated by positive learning). If you were forced to build a brain that would implement mean-dispersion preferences, and would like higher mean but dislike ambiguity ("fear of the economic unknown"), you might do it by having the striatum and OFC-amygdala activate much as they actually seem to do.

[2] One might wonder why we care about separating theories that have the same observable implications about field data. The answer is that the theories will often make different predictions about responses to changes in the economic environment. Recall Friedman and Savage's justly celebrated example of the pool players who "act as if" they understand the laws of physics (probably from learning to play by trial-and-error). Whether they act as if, or truly understand, physical laws will make a difference when the friction of the surface is changed or the pool table is warped and uneven. Act-as-ifers will make mistakes in the new environments but physical-law-knowers won't. Since economic environments are also undergoing constant change, the "as if" vs "really do know" distinction is important in economics too.

Is this account "too speculative"? It *is* speculative, but at this point speculating is all we can do and is necessary to make progress. Speculation of this sort is simply a process of conjecturing interpretations in a way that can be tested in more experiments.

Another potential use of neural data is to speed up the process of resolving fundamental debates. For example, Harrison (2008, this issue) has suggested that "we have conceptual work to do before we fire up the scanner", in the context of the debate about whether ultimatum-game rejections are due to a taste for reciprocity built into utilities, or reflect "field-hardened heuristics for playing repeated games". His view seems to be that we shouldn't do any brain scanning until the "conceptual work" is done.

However, these two interpretations about ultimatum rejections (as well as other prosocial behaviour like public goods giving) have been around for decades. There is no resolution in sight. Therefore, it is possible that the resolution could come much faster if we "fire up the scanner" soon rather than waiting for the conceptual work project to be finished. For example, if there was some agreement on what brain activity would result when a person who has developed "field-hardened heuristics" adapted to one environment is placed in an environment where those are no longer adaptive, or what it would mean neurally to have a utility for reciprocity in one-shot games, we could do the experiments and see what happens.

Mechanisms suggesting "new" variables: The most exciting possibility is that the neural studies will give us an empirically disciplined way to talk about variables that are thought to be important but which do not fit neatly into the fruitful preferences-beliefs-constraints structure (as in Spiegler's (2008, this issue) example of indecision).

For example, Rubinstein writes that "once we have enriched models of bounded rationality, it would make sense not to simply invent procedures from off the top of our heads but to use models based on our understanding of the mind".

There are many other potential examples: The nature of brain plasticity and human capital formation; the role of emotions and direct social interaction in service businesses; implicit and explicit discrimination in labour markets; motivations of workers and how business practices harness those motivations; genetic and intergenerational transmission of preferences and beliefs in families (with implications for demographic structure, and management of family firms); changes in neural activity across the life-cycle; behavioural genetics as a basis for understanding individual differences; emotional self-regulation in decision making; the nature of social interaction and emotion in service purchases in different media; the neural basis of contagion and panics in macroeconomic crises, and so on.

EPISTEMOLOGY AND EMPIRICAL PRACTICE

Harrison rightly points attention to the epistemology of neuroeconomics – the basis for knowledge claims.

The standard attacks on the quality of the data produced are surprising and a little exasperating for at least three reasons:

(a) Neuroscientists worry about methods and data quality frequently and are constantly making improvements (and neuroeconomists even more so). Journals are filled with methodology articles and neuroscientists are constantly innovating in how different tools are combined, improving experimental designs and statistical efficiency.
(b) Some people who do not know the methods work in detail are automatically suspicious. For example, Rubinstein (208, this issue) writes with typical candor that "colorful diagrams, which mean nothing to economists, are presented as clear evidence. To me, they look like a marketing gimmick . . . I almost always have the feeling of being (unintentionally) manipulated".

The fact that economists don't know how to interpret these diagrams is a statement about what economists don't know, not a statement about what neuroscientists do know. Fortunately, the knowledge gap will be narrowed as efforts are undertaken to communicate more across scientific language barriers (e.g. Glimcher *et al*., 2008).

With that said, it is true that a large degree of interpretive speculation is allowed in the field. Neuroscientists are expected and encouraged to speculate. Speculation is allowed because interesting results are quickly noticed and challenged by those who disagree with the speculation. This dialectic often results in clear conclusions rapidly (much more rapidly than in economics, which is almost glacial in comparison). Furthermore, the best diagrams usually offer clear evidence because results of many different studies are shown to activate common regions (in review journals like *Trends in Cognitive Science* and *Nature Reviews Neuroscience*). Economists should appreciate, therefore, that knowledge is reported and cumulated in two different ways – in "small" studies (typically with only one experiment) and in review articles that summarize dozens of those single-experiment studies, search for coherent themes, and sharply articulate open questions.

(c) Most dismissive criticisms ignore the fact that knowledge is produced by a *series* of studies with *different tools*.

For judging epistemology, the multiplicity of methods and cumulation across studies is absolutely a key point. Each method has limits, but limits are ideally compensated for by advantages of other methods. Furthermore, results of studies that are unclear are usually clarified by later studies.

fMRI is the most natural target for criticism because there is indeed something fascinating about seeing images of the brain. Harrison (2008, section 2.A) notes the many steps of "preprocessing" involved in matching up structural images (which have the best resolution of brain anatomy) to functional ones, correcting for errors resulting from head motion, and normalizing brains onto a common template. These steps are absolutely necessary if the goal is to gain power from across-subjects random-effects analysis.

Neuroscience has a good check on surprising results – rapid replication, and the requirement that hypotheses suggested by one method be corroborated by another method which has some advantages over the first one. Poldrack and Wagner (2004) suggest that "reverse inference" – the common practice of seeing that area A is activated differently in tasks T1 and T2 and then inferring something about activity in the tasks – should be considered simply a way to derive hypotheses which are tested in other experiments or with other methods. For example, the best studies find activity in interpretable regions and also see differences in activity in those regions which are correlated with behaviourally derived parameters across subjects.

In the Hsu *et al*. (2005) study, we used the fact that lateral OFC was more strongly activated in ambiguity tasks than in risk tasks to guess that patients with lesions to this region would be ambiguity-neutral. A separate study found that they were. Note that one can criticize the lesion study for its small sample ($N = 5$ lesion patients) and can criticize the fMRI study for its reverse inference speculation about the role of the OFC. However, one must criticize the general conclusion from the *conjunction* of the two studies since it is precisely their complementarity that matters.

Indeed, a great strength of neuroscience is that data come from many different methods – response times, psychophysiology (skin conductance, facial muscle recording), "single-unit" recording of neural firing rates, animal behavioural studies, genetic engineering, eyetracking, EEG, PET, fMRI, lesion studies, genetics, pharmacological manipulation, measurement of concentration of neurotransmitters, and TMS.[3] Each of these to some extent compensates for weakness of the others. For example, it is often difficult to know from an fMRI study whether a region has

[3] EEG is electroencephalography (recording from scalp electrodes, which is very fast and potentially portable), PET is positron emission tomography, and TMS is "transcranial magnetic stimulation", the use of a magnetic coil to disrupt activity in a brain area (sometimes called "temporary lesions"). Despite the ethical objections to TMS mentioned by Harrison, it is widely used and approved routinely by IRB boards that are charged specifically with making institution-level judgments about what is ethical and what is not.

been located that is causally involved in behaviour. But if that region is temporarily disrupted by TMS or permanently damaged by a lesion, and behaviour changes, then the two studies together (fMRI plus TMS or lesions plus TMS) have shown something more solid.

"STRAW MEN"

Harrison says that neuroeconomists "repeatedly set up straw men to knock down". I do not believe this is common at all.

A typical dictionary definition of a "straw man" is this: "To argue against a straw man is to interpret someone's position in an unfairly weak way, and so argue against a position that nobody holds, or is likely to hold."

For example, Harrison notes a passage in my 2004 paper about "state-contingent preferences" and says "Of course, economists have known this for decades". Then continuing, "Whether or not we use a state-dependent approach in analysis is a separate matter. The extreme alternative, *and no straw man*, is presented by Stigler and Becker (1977). They are clearly proposing the view that assuming that preferences are stable, and common across individuals, is simply a more useful approach [than state-dependent preferences]" (italics added).

So first we are criticized for endorsing state-contingent preference (as if the noncontingent approach is a straw man), then are told that the Stigler–Becker noncontingent approach is "no straw man". Which is it?

Economics is in the unusual position of having very simple specifications that are routinely used in analysis but which are widely known to be approximations that are easy to refute (e.g. expected value maximization). I do not regard these simplifications as straw men because they are *not* "positions that nobody holds" invented by a neuroeconomist; they are deliberate oversimplifications often used by economists as benchmarks. Furthermore, it is very useful to describe data in terms of deviations from these simplifications, as a way of learning which extended specifications are better.

A good example is self-interested preferences. This preference specification is clearly *not a straw man* because that simple form of preference is *actually used* in many types of analysis (e.g. in standard agency theory and empirical tests, in political science, and so forth).

Furthermore, belief in self-interest is not a position nobody holds because it is often clearly espoused. Stigler (1981: 176) wrote that when "self-interest and ethical values with wide verbal allegiance are in conflict, much of the time, most of the time in fact, self-interest theory . . . will win". List (2006) calls this statement "conventional wisdom among economists".

Even if economists don't believe it, self-interest is the routine assumption used in almost all most analyses (with important exceptions like the studies of bequests, intra-household behaviour, and charitable giving, where narrow self-interest is obviously violated). Using self-interest as a benchmark assumption is simply a way to learn which of these types of preferences exist and studying their nature.

FINALLY, LANGUAGE

Glenn Harrison is one of those people who is known for his remarkable skill in choosing language to infuriate even people who largely agree with him. (This is not an ad hominem attack; it is an empirical fact.) In his paper he refers to aspects of neuroeconomics as "academic marketing hype", "fluff", (and in another paper "jingoism"!?). I don't understand what these dismissive terms mean or how they apply.

For example, in one of my papers we describe the problem of how to model "choosing" to fall asleep at the wheel while driving. We jokingly say that one could think of such an action as revealing a stronger preference for sleep than for life. Glenn complains that "this is lousy, sucker-punch economics" and "can be waved away by any student of economics". Our satirical "explanation" is indeed "lousy economics" – that is exactly the point.

But what is the correct unlousy economic analysis of the sleepy driver's choice? Our point is simple (and we thought, obvious): You could conceivably describe this choice in the language of preferences, beliefs or constraints but it would be silly to do so, and it would tell you very little about designing cars or which public policies might work (such as offering free coffee on holiday weekends to tired drivers). For example, in his rejoinder Harrison says "any graduate student" would say that "the answer is that the driver chose a lottery when he got in the car tired, and his subjective probabilities of the various outcomes might have differed from the actuarial probabilities". But as the sleepy driver gets sleepier, the gap between subjective and actuarial probability widens. Moments before actually falling asleep the driver thinks there is no chance of an accident and the actual chance is quite high. Why does the gap widen? One can describe the mechanism in terms of subjective and objective (actuarial) beliefs but the entire point of the example is the dynamics over time.

What is wrong with inferring that the level of fatigue should be considered a state variable which is partly exogenous and partly under conscious control, and the fatigue level itself affects some kinds of motor control and judgment which affects "choice". That is, fatigue both creates a gap between perceived and actual risks, and influences the ability of the perceptual system to close the gap. This can certainly be described in the language of economics but it requires an admission that a visceral state both influences belief and influences processing of information in updating belief. A similar Chicago-esque view is that there is scarce "capital" in the form of attention or cognitive control, which is needed to make optimal decisions while driving, and fatigue temporarily lessens that capital. Fatigue is only one such visceral variable – others which might be important in economic and political decision making include lust (ask Elliot Spitzer and Bill Clinton), inebriation, pain, anger, hunger, fear and so forth. Neuroscience tools will help us understand how all these processes and emotions work in the brain (as will evolutionary psychology and many other fields).

Harrison's paper also includes some vaguer broad claims. Note that neuroscientists and neuroeconomists have many points of view and ways of doing and reporting their research. Simply because one paper may lack technical detail (from Glenn's perspective) should not indict an entire field. The fact that Marxist economists can join the AEA and come to its annual meetings does not mean that the economics profession should be criticized for "its" Marxism.

For a blog or casual conversation it is okay to use language as he does, but in a scholarly journal these claims require citation. Some of them are serious accusations. I challenge him to cite specific examples.

"In many cases it is impossible to figure out, without an extraordinary amount of careful reading, exactly what was done in the economics part of the experiment and statistical analysis." How many cases is "many"? Name several. [In his rejoinder, Harrison names four articles; I regard them as clear enough but readers can judge for themselves. He also does not answer the question of how many is "many".]

"One wonders how incestuous and unquestioning the refereeing process has become in certain journals". Name the "certain journals". What evidence or inference makes you think the process is "incestuous and unquestioning"? [In his rejoinder, Harrison names *Nature* and *Science*, but does not give any evidence or describe his inference except to assert that they have "low standards for science".]

". . . some of the debates on discounting behaviour, risk aversion 'calibration', and the existence of loss aversion, have taken on a resolutely thuggish tone that discourages useful discussion." Give an example of a "thuggish tone". Explain why it discourages useful discussion. [In his rejoinder Harrison does not answer these questions.]

"From the perspective of economists, the neuroeconomics literature seems to have used a production function with a sub-optimal mix of human capital and physical capital". What is the right mix? Which authors are not supplying enough human capital? Whose human capital should be used? [In his rejoinder, Harrison says "too little of the former [human capital]" but does not say which authors are supplying too little or whose human capital should be used.]

"As the behavioural economics literature demonstrated, however, we already knew how to do poor economics (and get it published)". Give an example. If the papers you name are so "poor" how did they get published? [In his rejoinder, Harrison does not give an example from behavioural economics. His answer to "how did they get published?" is "poor refereeing, or low standards for science".]

"Since economists have important and serious questions to get on with, the opportunity cost of these diversions [academic marketing claims] has just become too great to ignore." What are the important and serious questions? How do these "diversions" incur an opportunity cost? Why is the opportunity cost "too great to ignore"? [In his rejoinder, Harrison does not answer these questions.]

REFERENCES

Colander, D. 2008. Neuroeconomics, the hedonimeter, and utility: some historical links. *Journal of Economic Perspectives*.

Fox, C. and R. Poldrack (in press). Prospect theory in the brain. In *Decision neuroscience*, ed. P. Glimcher, C. Camerer, E. Fehr and R. Poldrack. Amsterdam: Elsevier.

Grant, S. and B. Polak. 2007. Absolute ambiguity aversion: Mean dispersion preferences. Yale working paper.

Glimcher, P., C. F. Camerer, E. Fehr and R. Poldrack. 2008. *Decision neuroscience*. Amsterdam: Elsevier.

Harrison, G. 2008. Neuroeconomics: A critical reconsideration. *Economics and Philosophy* 24.

Hsu, M., M. Bhatt, R. Adolphs, D. Tranel and C. F. Camerer 2005. Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310: 1680–3.

List, J. 2006. When the behavioralist meets the market. *Journal of Political Economy*.

Poldrack, R. A. & A. D. Wagner 2004. What can neuroimaging tell us about the mind? Insights from prefrontal cortex. *Current Directions in Psychological Science* 13: 177–81.

Rubinstein, A. Comments on neuroeconomics. *Economics and Philosophy*.

Spiegler, R. 2008. Comments on the potential significance of neuroeconomics for economic theory. *Economics and Philosophy* 24.

Stigler, G. 1981. Economics or ethics? In *The Tanner Lectures on Human Values*, Vol. 2, ed. S. McMurrin. Cambridge: Cambridge University Press.