Economics and Philosophy, **24** (2008) 345–368 Copyright © Cambridge University Press
doi:10.1017/S0266267108002010

NEUROECONOMICS AND THE ECONOMIC SCIENCES

KEVIN A. MCCABE
George Mason University

Neuroeconomics is the newest of the economic sciences with a focus on how the embodied human brain interacts with its institutional and social environment to make economic decisions. This paper presents an overview of neuroeconomics methods and reviews a number of results in this emerging field of study.

INTRODUCTION

This essay is a brief introduction to neuroeconomics with the goal of answering the following question: Is neuroeconomics useful for the study of economics? The short answer is yes, but the contribution so far, based on published research, is small. This is partly because neuroeconomics is less than ten years old, and partly because the conceptualization of economic behaviour in terms of neuronal computation is difficult. But, the contribution of neuroeconomics is likely to grow because of the successes of its parent disciplines: economic science and neuroscience. In particular, the usefulness of neuroeconomics to economics research is tied to the broader question of the usefulness of the experimental method in economics; a question which many believe is largely answered, and if so, it only remains to show how neuroeconomics experiments can complement already existing experimental methods. The usefulness of neuroeconomics is also tied to our knowledge of how the brain functions, and how we measure brain function. Current knowledge is incomplete, but proceeding at a rapid pace, and soon we will have complete-enough models of brain function *to help understand* and solve important economic problems.

The goal of this essay is not to answer the question, is neuroeconomics important? Or even the question, is neuroeconomics important to the discipline of economics? These questions are easier to answer because the experiments designed in economics laboratories have shown themselves to be very useful for the study of humans (and other non-human species) in neuroscientists' laboratories. Furthermore, the theories of subjective preferences, optimal decision making, and decentralized coordination have been found useful in talking about the organization of the brain. The human condition is intertwined with economic decisions, and institutions, and neuroscientists interested in decision making and social interaction are interested in studying the normal, and impaired, brain as it engages in economic activities. It is not surprising therefore that neuroeconomics is important in this dialogue. Economists should be interested in this development since neuroeconomics is becoming an increasingly important route for the export of economic ideas. What remains is to ask: how important is the study of the brain to the study of economics?

To answer this question we have to ask what is the study of the brain? There is no simple answer, but it may help to consider the same question posed by cognitive psychology in its evaluation of the contribution of cognitive neuropsychology. Cognitive psychology models cognition in terms of information processing done in the brain. A common, noninvasive, means to study human brain activity is with PET (Positron Emission Tomography), and fMRI (functional Magnetic Resonance Imaging). A critique of using imaging experiments in cognitive psychology was put forward by Harley (2004) and Coltheart (2004), who argue that studying brain images can tell us nothing about information processing. For example,

> The assertion is that this aim, *i.e., the aim of using neuroimaging data to test between cognitive models*, is impossible to achieve in principle, because facts about the brain do not constrain the possible nature of mental information-processing systems. (Coltheart 2004: 22; italics added).

The idea is that the brain and the mind should be treated as two independent entities much like computer hardware and the software instructions that run the hardware. Thus studying what the hardware is doing is useless because it doesn't tell us about the algorithm that is implemented in the software. But, the evolutionary account of the brain shows that the brain evolved to satisfy specific functions; this suggests that the hardware/software dichotomy is a bad analogy.

The evolution of the human brain over the last two million years has been remarkable. The most obvious fact is the brain has more than doubled in size even though the energetic cost of an ounce of brain is five times that of an ounce of muscle; from this we might infer that the human ecological niche depends more on brain than on brawn. The brain consumes roughly 20 % of the human body's energy, and yet, it runs over a hundred billion neurons with less energy than it takes to run a hundred watt light bulb; see Montague (2006) for a more complete account of the implications of scarcity on brain adaptation. While many different neuronal systems communicate with each other in parallel the speed of this communication runs at the relatively sedate pace of a hundred metres per second. Thus the right analogy is that the brain runs proprietary neural assemblies, designed by evolution and updated by experience, to maximize specific functional capabilities given the physical constraints of the system. Furthermore, as a theory of optimization under scarcity would predict, these systems are located where the relative supply of the right kinds of neurons are plentiful and where their ability to service demand is high.

Given a systematic mapping between brain function and brain structure, imaging data becomes simply another dependent variable that can be used to test between theories (Henson 2005). For example, Henson *et al.* (1999) use an imaging experiment to test a dual process model of recognition memory against a single process model. What is required is that within an experimental design it is not the case that one set of brain regions are active with a cognitive function in one condition, but that another set of brain regions are active for the same cognitive function in another condition.

Houser *et al.* (2007) has made the case more succinctly for neuroeconomics. The canonical economic problem is to make inferences about the probability, $p(d_{t+1} \mid x_t)$ that someone will make a future decision, d_{t+1}, given their contemporaneously observable individual characteristics, x_t. If, in addition, each person can be identified by a latent variable k that can represent unobservable preferences or cognitive strategies, such that $p(d_{t+1|}\ k, x_t)$ holds, then information from behavioural and brain imaging experiments can help to identify k probabilistically as $p(k \mid x_t)$. From this formulation we can then derive,

$$P(d_{t+1},\ x_t) = \Sigma_k\, p(d_{t+1}|k,\ x_t)p(k|x_t).$$

The authors go on to show in a model with two possible types of conditions for which imaging data can be combined with behavioural data to get better estimates of $p(k \mid x_t)$.

We can compare the quote by Coltheart above to a similar critique by Gul and Pesendorfer (2005) only now used to critique neuroeconomics.

> Rationality in economics is not tied to physiological causes of behavior and therefore the physiological mechanism cannot shed light on whether a choice is rational or not in the sense economists use the term. (Gul and Pesendorfer 2005: 24)

Note, in their critique, Gul and Pessendorfer are trying to avoid the hardware/software analogy by saying something stronger. The software they are interested in does not run on any particular computer but instead is some ideal software running on an ideal virtual computer. It is clearly not the human brain. Why this stronger form? In part it may be the fear that we already have too many degrees of freedom to model economic behaviour. Given that preferences can be any weak ordering of alternatives and strategies can be any mapping from a partition of information to actions it turns out that economics can already explain just about any kind of behaviour. For example, Ledyard (1986) has shown that almost any profile of strategies can be as Bayesian equilibrium of some game. The fear may be that allowing for cognitive heterogeneity can only make the theory more vacuous. However the opposite is more likely to be true. By moving the study of decision making a step down, to the brain, neuroeconomics will put more structure on preferences, encoding of information, and cognition, and thus will make economic theory more predictive and consequently more useful.

As this essay was being written the Society for Neuroeconomics met in New York to announce the forthcoming publication of the *Handbook of Neuroeconomics*.[1] While neuroeconomics is only ten years old the conference attracted over 600 people. Why is this so? The answer lies in part with the fascination that humans have with the human brain. We all suspect, even if we don't state it, that understanding the human brain is the key to following the sage's advice to "know thyself". For many centuries economists have used reason, informed by observation of the external world, to build models of human behaviour, and by doing so we have learned a great deal. But, as we move into the twenty-first century, we will have an increasing ability to see inside our brains and understand how brain activity produces economic decisions. It is certain that our observations of the internal world of the brain will also be used to inform our theories of human behaviour.

An important problem of this century is to understand the disparity of economic growth, and material welfare, both between and within nations. This is not a problem that any single discipline can successfully address, but many disciplines can play a role. Neuroeconomics is no exception. Douglass North (2005) makes a good case for the importance of understanding how our brains form beliefs over which we act, and how these actions shape the culture that we live in. Neuroeconomics provides us with some answers, even if they are still provisional answers. The goal-directed brain is designed to use reinforcement learning strategies to optimally use the impersonal rules defined by an institution, but the same brain is also designed to use reciprocity to build personal networks that allow us to sometimes circumvent, and sometimes change the rules of our institutions. In understanding human exchange behaviour we had to understand how this conflict gets resolved in our brains. Only then can

---

[1] The interested reader might find the handbook very useful to learn about the many interesting neuroeconomics projects only some of which have been discussed in this essay. More can be found out about the society and the handbook at www.neuroeconomics.org.

we begin to understand how the incentives and information produced by our institutions resolves this conflict in order to balance the use of personal and impersonal exchange in a civil society; a balance that has a dramatic impact on economic success.

WHAT IS NEUROECONOMICS?

Collaborative research between cognitive neuroscientists and economists is producing a new interdisciplinary field called Neuroeconomics (McCabe 2002; Glimcher and Rustichini 2004; Camerer *et al.* 2005). Neuroeconomics begins with the observation that humans face many opportunity cost tradeoffs in their daily activities. These tradeoffs are modelled as a strategy that maps information sets into actions, and as is assumed in evolutionary game theory (see Maynard-Smith 1982; Weibull, 1996) the evolutionary process will select strategies that maximize inclusive fitness. Herbert Simon's research on bounded rationality, (Simon 1957) implies that strategies are likely to be encoded in the brain as a mapping from partitions of circumstances into partitions of actions together with inferential (Holland *et al.* 1986) and reasoning mechanisms (Gigerenzer and Selten 2001) that modify and scale these partitions. To understand how such encodings and mechanisms are formed requires both a top-down approach using experimental methods and strategic models from economics and a bottom-up approach using experimental methods and computational models from cognitive neuroscience.

THE NATURE OF NEUROECONOMICS COLLABORATIONS

Economists and neuroscientists are both interested in performance, but not the same performance. Economists are interested in markets and how the rules of the game embodied in institutions affect market performance. Economists are also interested in preferences and choice theories, but as ways to understand the individual in order to study markets. On the other hand, neuroscientists are interested in neurons and how collections of neurons assemble and communicate to affect cognitive performance. Neuroscientists are also interested in markets, but as natural places to study cognition. Neuroeconomics reconciles these interests, first by putting more structure on preferences and the information processing that supports decision making which allows us to say more (not less) about market performance, and second, by recognizing that institutions make computations that extend our cognitive performance.

Neuroeconomics experiments use a combination of brain imaging/stimulation experiments developed in the cognitive neurosciences and microeconomic systems/game theory experiments developed in the economic sciences. The fact that both disciplines shared similarities in doing experiment made it possible for neuroeconomics to emerge as

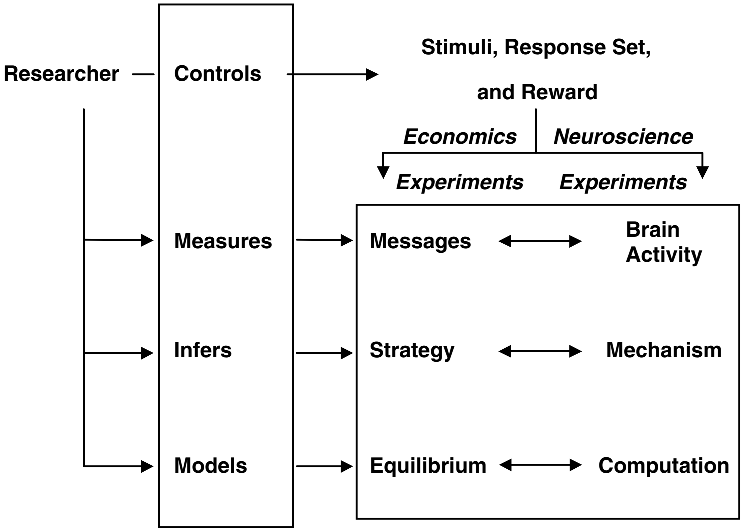

FIGURE 1. Economics and cognitive neuroscience experiments compared.

economists worked with neuroscientists to explore problems of mutual interest.

From Figure 1 we see that researchers in both disciplines try to control for the stimuli that subjects see, the response set available to subjects, and the reward that subjects receive for their responses. However, what the two disciplines measure in their experiments, and the models they used to design their experiments, are very different, but it turns out these differences can work together to provide interesting neuroeconomics experiments. At the risk of oversimplifying, economists measure messages, while neuroscientists measure brain activity; but, brain activity is what produces the messages providing an additional source of data to characterize subjects. Second, economists try to infer subjects' strategies from the messages produced, while neuroscientists try to infer the neuronal mechanisms that produced the observed neural activations; but strategies are representations of the function that the mechanisms are performing. Third, economists model the observed strategies in terms of equilibrium conditions, while neuroscientists model the neuronal mechanisms in terms of the computations they perform; but equilibrium and computation are organized by the general principle of optimization. What has emerged is a generally congruent model of human behaviour shown in Figure 2 and explained in the next section.

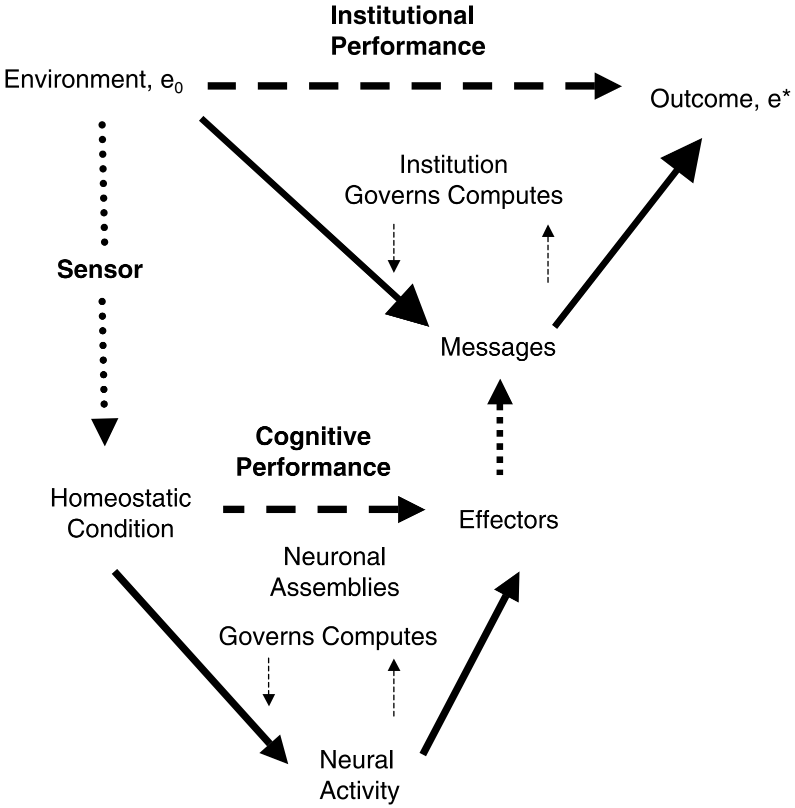

FIGURE 2. Microeconomic and Neural Systems.

RESEARCH METHODS

A microeconomic system (see Smith 1982) is an ordered pair (e_0, I) where, $e_0 \in E$, is a particular instance of an economic environment and I is a particular instance of an economic institution. As shown in Figure 2 (top half), an experiment is designed to capture the initial conditions as specified by the environment, and the institutional rules that allow the agents to change their environment. The environment consists of agent descriptions, including agents' preferences, their technological know-how, and their initial holding of goods and other resources. The institution consists of a message space, M, together with governance, and production, rules. Governance rules determine what messages agents are allowed to send, and when they are allowed to send them. Production rules determine how the environment changes as a function of the messages sent. Game theory is a special case of a microeconomic system where the environment and institution are specified by the von-Neumann Morgenstern preferences over strategy profiles.

In designing an economics experiment the experimenter maps subjects to agent descriptions. A critical methodological component of this mapping is the induced value procedure (see Smith 1976) which pays subjects a salient, usually monetary, reward that depends on the end state, e^*, achieved by subjects in the experiment. This procedure is important for maintaining experimental control and allowing more consistent replication. Another feature of many economics experiments is an explicit performance function, $F: E \to R$, which allows the experimenter to compare different economic institutions in the same environment, or the effect of a changing environment on an economic institution. Given a performance function, F, economists are often less interested in understanding agents' behavioural strategies, defined by $b_i: E \times I \to M$, at least as long as the theory is working. However, when the theory fails to predict the end state of the experiment with reasonable accuracy the study of subjects' behaviour strategies becomes important in order to develop insights into how to change the theory. It is at this point that neuroeconomics experiments are valuable.

A neuronal system (see LeDoux 2002), shown in the bottom half of Figure 2, consists of an embodied brain that is trying to maintain homeostatic equilibrium while responding to environmental opportunities. Changes in homeostatic condition and/or opportunities result in neuronal activity which is governed by the synaptic connectivity of neuronal assemblies and may ultimately result in motor responses that map onto the message space. The synaptic weights change over time based on the firing of a connected train of neurons that process internal (homeostatic) information and/or external (sensory) information as it makes its way through the nervous system. Conceptually, a neural system is very similar to a microeconomic system, but the process of designing an experiment to study neural function is quite different since a neural system does not define an agent to which a subject can be mapped.

In designing a cognitive neuroscience experiment the experimenter must decide whether or not to study human or nonhuman brains. The tradeoff is obvious. The study of human brains, except in relatively rare medical cases, must be non-invasive and non-injurious to the subject thus constraining the types of experiments that can be run. As a consequence nonhuman brains are often used as models of the human brain based on the evolutionary hypothesis that the human brain evolved over time by extending homologous functions, and computations, in predecessor brains (see Gazzaniga *et al*. 2002). In nonhuman brains the experimenter can use lesion experiments, genetic manipulations, neuronal stimulation of specific neurons, and single/population cell (as well as extra cellular) recordings to measure neuronal signals, and neuronal connectivity, in order to reconstruct neuronal mechanisms and learn how they compute (see Gazzaniga *et al*. 2002). Without this basic research it would be difficult to understand the data from experiments on human brains.

In studying the human brain the neuroscientist also has the option of studying brains that have suffered brain injuries or illnesses. The injury and illness studies allow the experimenter to observe the effects of abnormalities on neuronal activity and cognitive performance measured as the ability of a subject to maintain homeostasis and/or make goal improving responses. For the non-injured or non-ill brain that may also satisfy other constraints on the contemporaneously observable characteristics of the subjects (such as age, gender, mental performance measures etc.), herein called the "normal" brain, minimally invasive experimental methods are used. One of these, functional Magnetic Resonance Imaging (fMRI), is the most common in neuroeconomics experiments.

In an fMRI experiment (see Huettel *et al*. 2004), the Blood Oxygen Level Dependent Response (or BOLD response) is measured by detecting changes in magnetic resonance signals produced by spinning protons in a strong magnetic field. Signal changes allow the experimentalist to measure the blood oxygen levels of capillary fields that supply nearby neurons. Recent experiments have begun to explore the connection between single cell firing in primate brains and their BOLD response as measured by fMRI (see Logothetis *et al*. 2001) allowing us to better understand human BOLD response to the underlying neuronal mechanisms being discovered in non-human experiments.

In designing an fMRI experiment (see Huettel *et al*. 2004) the experimentalist chooses a stimulus-response-reward design that is most likely to produce a measurable BOLD response and then measures the magnitude of the response relative to when (what) stimulus occurs, when (what) action is taken, and when (what) reward or outcome occurs. Many designs involve the subtraction method where the experimentalist compares BOLD differences in two (or more) slightly different stimulus-response-reward treatments in order to explore the neuronal basis of the hypothesized computational (or functional) differences involved. Data are subsequently analysed and hypotheses tested using the General Linear Model on the spatial and temporal dataset. While current practice is relatively crude, advances are being made in the technologies for detecting the MR signal, the kinds and properties of neuronal systems that can be measured, and the statistical methods employed.

WHAT DO NEUROECONOMISTS DO?

To understand the role of neuroeconomics in the economic sciences it helps to understand what neuroeconomists do and put this into context with what economists are interested in. A popular textbook definition of economics is that it is the study of the allocation of scarce resources to unlimited human wants. This definition suggests that economists should be interested in the nature of human wants, how they are expressed, and how the allocation of scarce resources satisfies these wants.

This section is broken down into four subsections. The first section defines wants in terms of goal-directed behaviours. Brains are well designed for goal-directed behaviour using reinforcement learning mechanisms and a fair amount is known about how the brain computations can produce such behaviour (see Sutton and Barto 1998). Moreover, humans invent institutions, such as money, to facilitate our goal-directed behaviours and these institutions in turn get instantiated in the brain. The second subsection studies how goal-directed behaviours interact with cognitive control mechanisms to produce decisions of particular interest to economists. The third and fourth subsections study the mechanism of exchange in optimizing our goal-directed behaviours. The third subsection looks at personal exchange and how the principles of reciprocity are instantiated in the goal directed brain. The fourth subsection looks at impersonal exchange and how the price system is used in the brain to facilitate goal-directed learning.[2]

The nature of wants and goal-directed behaviour

Decision-making requires the coordinated activity of motivational, emotional and cognitive circuitry to plan, discern and weigh alternatives, take actions, and learn from appropriate feedback. It seems that all animals, including humans, have an instinctive desire to explore their environment; this in turn requires neural mechanisms that (a) detect rewards and punishments, and (b) learn policies that map state representations onto a distribution of actions. As a consequence, goal directed animals learn to approach and produce states associated with rewards while avoiding and preventing states associated with punishments.

We can start with the following simple thought experiment. You are thirsty, and as you are walking to your office you remember your colleagues are having a conference lunch and often leave leftover drinks for anyone who wants them. At this point you have a decision to make: do you walk directly to your office, or take a short detour to check on the possibility of there being leftovers? If the odds are good enough, and you have the time, you probably will take the detour.

This decision seems to be controlled by two brain systems (Balleine *et al.* 1998). The first system is the stimulus-response/reinforcement system

---

[2] See also Singer (2008) and Houser (2008).

proposed by Thorndyke (1911), which encodes the correlation between an action and a reward/penalty. The second system is a "goal directed" system which (a) identifies a contingent relationship between one's action and a resulting reward/penalty, and (b) allows the individual to evaluate different responses in terms of the relevance of current or anticipated motivational states.

These systems evolved over time in the mammalian brain; for example, it is well known that neuronal ensembles for both systems are instantiated in the rat brain (White 1989; Connover and Shizgal 2005). The rat's prelimbic region, which later becomes the dorsolateral prefrontal cortex in monkeys and humans, seems to be essential for the operation of the contingent-learning system of the causal relationship between action and reward. A separate system in the insular cortex seems to be important for connecting motivational states to outcomes.

Similar systems are also found in the monkey brain. The motivational value of rewards as processed by areas of the brain connected by the dopaminergic pathways starting with the production of dopamine in the ventral tegmental area, and then projects to the striatum (caudate nucleus and putamen), the ventral striatum (nucleus accumbens), and then outwards to neocortex (including anterior cingulate cortex, and orbitofrontal regions of the brain) (see review by Schultz 2000).

Schultz *et al.* (1997) argue that the activation of midbrain dopamine neurons in monkeys function to produce goal-directed learning; in particular, midbrain dopamine neurons show a burst of activity when a novel reward occurs. Once the reward becomes associated with a stimulus this burst of activity shifts from the time of the reward to when the stimulus first appears. Thus the neurons shift their activity to the anticipation, or prediction, of reward. Yet, if the stimulus occurs and the reward is not delivered when expected, the same midbrain dopamine neurons decrease their rate of activation at the time the expected reward should have occurred. This suggests that the dopamine system reports discrepancies between the prediction of a reward and the occurrence of a reward, called the temporal difference error, and that such an error signal would be essential for learning. It is this anticipation of reward that most closely maps on to the economist's notion of wants.

O'Doherty *et al.* (2004) find equivalent regions of BOLD response in the human brain when subjects either received, or made choices, to get juice rewards. They find that the ventral striatum is active as the subjects wait in anticipation of getting a risky juice reward, while the dorsal striatum is active when the subjects choose an action to get a juice reward. These areas coincide nicely with the actor (dorsal striatum) – critic (ventral striatum) model of reinforcement learning also found in monkeys, and is consistent with the view that this system enables the goal-directed learning of strategies to maximize intrinsic returns; as such it suggests an interpretation of BOLD responses in these regions as a measurement of a subjects' wants.

In summary, we can now answer the question "what do people want?" From a biological perspective, wants are derived from the end goal of maximizing inclusive fitness, and are expressed through the more proximal goals of maintaining homeostatic equilibrium and maximizing reproductive success. From a neuroscience perspective these goals are achieved through the instantiation of motivational value based system that drive goal-directed learning. Therefore, what people want is to maximize the return on their goal directed behaviours.

Now consider the following thought experiment. You are walking along and you see a dollar bill. Do you pick it up? Most people would say yes. Note that the dollar has benefits only if you spend it, thus making it different from a drink which benefits you immediately. However, in both cases, most people would see picking up the dollar as also satisfying your immediate self-interest. So, how is the motivational value for picking up the dollar encoded in the brain?

As it turns out, the motivational value for the dollar bill is instantiated in the same part of the brain as the juice rewards. Knutson *et al.* (2000, 2001a, 2001b) design an experiment where human subjects first see a cue indicating how much money they could make, and then after some delay they see a square which appears for a brief amount of time. If a subject presses a button while the square is present, the subject earns the monetary amount given by the cue. The authors find that the right nucleus accumbens in the ventral striatum is active during the delay, when the subject is anticipating earning the money, and the right ventral medial prefrontal cortex is active when the subject receives the reward. Furthermore, they find that the level of BOLD activation in both areas varies with the size of the reward suggesting that dopamine neurons encode the saliency of the reward.

One conclusion that can be made from these, and similar studies, is that the brain economizes on scarce neuronal resources. In outcomes that produce both direct functional rewards, such as juice, and indirect functional rewards, such as money, the motivational system encodes an anticipated reward value for the outcome in order to invoke an action. When the outcome is associated with the action, the motivational system calculates a realized value, compares this to the anticipated value, and only when there is a difference does the brain commit scarce resources to update its strategy. An important lesson to learn is that the brain is capable of intrinsically valuing an institutional or cultural construct, such as money, based on its end use, and then using this valuation to motivate learning.

McClure *et al.* (2004a) studies how circuits in the brain process a cultural construct, in this case product brand names, in order to affect behavioural choices. In their study subjects were given either a blind taste test of two soft drinks, Coke® and Pepsi®, and asked which they prefer or they were given a taste test over an unlabelled cup of soft drink, which they were told could contain either brand but in fact always corresponded to the soft drink in the labelled cup, or a labelled cup containing that brand of drink. In the anonymous taste task subjects were equally likely to say they preferred Pepsi® to Coke®; a choice which often disagreed with their claimed preference. However, in the labelled cup choice subjects were much more likely to say they prefer the cup labelled Coke®, compared with the Coke® in the unlabelled cup, even though the only difference was the label. However, subjects were still equally likely to say they prefer the cup labelled Pepsi®, compared with the Pepsi® in the unlabelled cup. This suggests that the Coke® brand name increased the likelihood of choosing that cup.

McClure and his colleagues then scanned the brains of subjects who in one session saw a coloured light (one for each soft drink) followed by the delivery of a soft drink through a plastic tube (while they were in the scanner.) In a second session, with different subjects, the Coke® light was replaced with a picture of a Coke® can. In the first session they found that subjects' ventral medial prefrontal cortex monotonically increased in activation when they received the drink (monotonic to the number of times the subject chose that drink in the anonymous taste task). This activation is very similar to the activation Knutson *et al.* found with the realization of monetary rewards. However, in the second session, when subjects saw the Coke® can additional areas in the brain were activated including the hippocampus, dorsal lateral prefrontal cortex, and the midbrain, leading the authors to conclude that these areas might operate to bias preferences based on cultural (brand name) information.

Economic decisions and goal-directed behaviour

Time and uncertainty play a major role in economic studies of market behaviour. Preliminary evidence suggests that choices in such environments use a number of neural systems either working together or in opposition to each other.

Uncertainty

Kuhnen and Knutson (2005), study subjects choices over two risky and one riskless gambles (A, B, C), with A returning ([\$10, 0.5], [\$0, 0.25], [−\$10, 0.25]), B returning ([\$10, 0.25], [\$0, 0.25], [−\$10, 0.50]), and C returning ([\$1, 1.0]), where [\$x, p] specifies that \$x will be won or lost (if negative) with probability p. Notice that A is better than B in terms of first order stochastic dominance. Subjects made choices in 20 blocks, each consisting of ten trials, having four stages; an anticipation stage (they see the choice problem); a choice stage (when they make a choice); an outcome stage (when they see the outcome for the asset they choose); and, a market stage (where they see the outcome for all the assets). The authors found that choices among gambles were correlated with BOLD responses in the nucleus accumbens and the anterior insula, and that activations in these areas can be used (in logistic regressions) to predict the probability of making subsequent choices. They conclude that differential activation in these regions may lead to either risk-seeking or risk-averse behaviour.

In a separate study, Preuschoff *et al.* (2006) study the BOLD response in the ventral striatum in repeated trials of a gambling task, where they systematically varied expected reward and risk (variance of reward). In their task a subject, initially endowed with $25, must place a bet of $1 on whether a first card to be drawn (from 1 to 10 with replacement) will be higher then a second card drawn (from 1 to 10 with replacement). After they make their bet, the first card is then shown, followed 7.5 seconds later by the second card. The subject earned $1 if they were correct, and lost $1 if they were incorrect; as expected subjects won roughly 50 % of the time. Note that subjects could update their probability of winning, and thus the riskiness of the bet, after seeing the first card. At the time the first card was displayed the ventral striatum of subjects showed BOLD responses that correlated linearly with expected reward. However, in the time period leading up to the revelation of the second card BOLD responses in the ventral striatum showed an activation pattern consistent with the quadratic encoding of the updated probability of winning. The authors interpret these results as showing that the ventral striatum encodes both risk and expected return.

How does the brain choose when faced with conflicting representation of the goal to be achieved? Conflict-resolution is known to include areas of prefrontal cortex and the anterior cingulate and has been studied experimentally and modelled formally by Cohen and his coauthors; see the review by Botvinick *et al.* (2000). One example of conflict monitoring and conflict resolution is the well known Stroop task. In this task there are words for different colours, such as RED and BLUE, which are presented to subjects in different coloured letters. For example the word RED may be presented with blue letters. Subjects are then asked to say as quickly as possible either the colour that the word names (the word task) or the colour of the letters the word is composed of (the colour task.) For example, suppose you see the word RED in blue letters and are charged with doing the colour task. The right answer is blue, but the tendency to want to say red is very strong. In this case the prepotent response, i.e. red, is wrong, but the conflict resolution mechanism not only produces the correct response, but weakens the prepotent response in the future making it easier to solve the colour task next time.

Conflict resolution mechanisms have been used to explain hyperbolic discounting in making intertemporal choices. McClure *et al*. (2004b) study BOLD activations in subjects who were required to make choices between an immediate monetary reward and a distant reward, or to make choices between a distant reward and an even more distant reward. They found that choices that included immediate rewards were correlated with BOLD responses in the limbic system including the ventral striatum, medial prefrontal cortex, and posterior cingulate cortex. By comparison all choices, i.e. those including an immediate option or those including only distant options, showed activation in the lateral prefrontal areas and associated parietal areas. Furthermore, when the immediate reward was available, but subjects chose the delayed reward the authors found greater BOLD responses in the lateral prefrontal and parietal areas of the brain compared to the limbic areas of the brain.

In a related study, Kable and Glimcher (2007) studied the choice between an immediate $20 reward and a larger delayed reward varying between $20.25 and $110.00. They then consider 10 subjects who displayed stable subjective discount functions that could be characterized by either a one parameter exponential discount function or a two parameter (the sum of two exponential functions) discount function. BOLD responses that were positively correlated with subjective values (based on subjects' fitted discount functions) were the medial prefrontal cortex, ventral striatum, and posterior cingulate cortex, the same limbic areas found in the previous study, but now surprisingly for choices including immediate rewards.

In conclusion, the Preuschoff *et al*. and Kabel and Glimcher studies provide some evidence that subjective decision variables found in standard economic models of choice over risky options or choice over temporal are encoded in the goal-formulation and goal-directed regions of the brain.

Strategic decision making and personal exchange

Economists use game theory to study strategic interaction. Of particular interest in this paper is how games are used to depict economic exchange. This section shows how neuroeconomics experiments have been used to complement laboratory experiments in the study of personal exchange. The first subsection considers a simple distribution problem (called the Dictator game) faced by a single individual between themselves and another person. It acts as a baseline in which to consider exchange. The second subsection considers a simple bargaining decision (the Ultimatum game) between two subjects. It studies the effect of threats (and negative reciprocity) on exchange. The third subsection considers a simple investment decision (the Investment game) between two subjects. It studies the effects of trust (and positive reciprocity) on exchange.

Dictator experiments

The "dictator game" was introduced under this name by Forsythe *et al*. (1994). In this experiment, a subject in room A is given $10 by the experimenter and is asked how they would like to split the $10 between themselves and an anonymous subject in room B. Their decision is final. The data breaks down as follows: Only 21 % of the subjects kept all the money, 17 % kept nine dollars, 12 % kept eight dollars, 29 % kept seven dollars, and finally only 21 % split the ten dollars evenly. Overall, 23 % of the money was sent to the person in room B.

Why do people send even 23 %? Hoffman *et al*. (1996) hypothesized that our evolved social brain would be sensitive to the likelihood of being seen as and/or found out to be non-cooperative based on one's group behaviour. They call the inverse of this likelihood "Social Distance". Thus the greater the social distance the less the likelihood of being typed as non-cooperative and the more likely one is to behave in one's immediate self-interest. In a double-blind experiment, designed to maximize social distance, the distribution is much more self-interested. Now, 64 % of the subjects kept all the money, 20 % kept $9, 7 % kept $7 or $8, and only 8 % split the money equally. Overall less than 9 % of the money was sent to room B.

In both of the experiments reported above subjects were given the $10 by the experimenter. Cherry *et al*. (2002) examine dictator giving when the dictator had to earned their money (in the experiment) before deciding how much to send. Using the double-blind control they found that dictators now kept the money 95 % of the time. These studies suggest that a subject's sense of ownership or right to the money affects how much they will give.

In a recent neuroeconomics experiment, Spitzer *et al*. (2007) study the BOLD responses of dictators when there is a likelihood that they could be punished compared to dictators who were safe from punishment. The authors find that areas involved in decision making, including the dorsal lateral prefrontal cortex, the ventral lateral prefrontal cortex, the lateral orbitofrontal cortex and the caudate, all showed greater BOLD responses in the brains of dictators who could be punished. It will be interesting to see how this changes when dictators earn their money.

The ultimatum game

The ultimatum game provides examples of both conciliatory behaviour by player 1, and negative reciprocity by player 2, while bargaining over the terms of a personal exchange. Player 1 must propose how to split a fixed amount of money, $10. Once player 1 proposes, player 2 can either accept or reject. If player 2 rejects the proposal, both players earn zero; otherwise, the players earn the split proposed by player 1. Rational choice theory predicts that player 2 should accept any positive offer, and player 1, reasoning this way, should offer player 2 some small amount; such as the proposal of ($9, $1).

When the ultimatum game was first run with cash-motivated subjects, Guth *et al.* (1982) observed that the modal proposal was to split the money 50-50. This result has been replicated dozens of times, including by Forsythe *et al.* (1994), who compare offers in ultimatum games to those in dictator games and show that the 50-50 proposals in ultimatum games are largely a consequence of player 2's ability to reject player 1's proposal. Thus, in order to reduce the risk of rejection, player 1 makes more conciliatory offers.

Hoffman *et al.* (1994) test the predictions of social exchange theory in the ultimatum game. Two social exchange conditions, (1) a contest in which subjects earn the right to be player 1 and (2) socially defined Seller/Buyer exchange roles for players 1 and 2, are compared to a baseline condition with (1) random assignment to the first mover position, and (2) neutral role definitions. In the baseline condition, half of the offers are at $5 with a mean offer to player 2 of $4.37. By comparison, the property right assignments with buyer-seller roles, which was predicted by Hoffman *et al.* to have the strongest equity norm effect, resulted in less than 10 % of the offers at 5-50 with a mean offer to player 2 of $3.08. In both cases, rejection rates were low, at about 10 %, suggesting that player 1's low offers were no more risky. This suggests that Player 2s implicitly recognized the right of their counterpart to offer less when they had earned the right to do so.

Neuroeconomics experiments have begun to study some of the neural underpinnings of bargaining in the ultimatum game. In a fMRI study, McCabe *et al.* (2001) study brain activation in 12 people who played sequential two-person simplified trust and ultimatum games. Half the time they were player 1, and the other half they were player 2. Each time they played, their counterpart was either a computer playing a fixed probabilistic strategy, or a person who was recruited to play outside the scanner. Subjects were told each play whether they were playing the computer or the person. The authors conjectured that subjects would use theory-of-mind (see Frith and Frith, 1999) regions of their brain (shown to exist in other studies) to infer the intentions of the other player and that this would play an important role in the binding of mutual payoff information to a cooperative event representation and thus invoke cognitively strategies for delay of gratification, and thus produce trust and reciprocity.

Based on their individual plays, five of the 12 subjects were labelled as cooperators while five were labelled as non-cooperators. In a conjunction analysis, the seven cooperators all showed greater prefrontal activations in the anterior paracingulate (known previously to be involved in theory-of-mind tasks) and dorsolateral prefrontal cortices (known to be involved in cognitive control tasks). The authors argue that the observed activation in cooperators is consistent with shared reciprocity intentions, resulting in both the inhibition of individual reward seeking by player 2, and the inhibition of risk avoiding behaviour by player 1.

Sanfey *et al.* (2003) use fMRI to study the neural correlates of the player 2 behaviour of 19 subjects in the ultimatum game. In their experiment, subjects in the role of player 2 made 20 decisions. Ten games were played with other people and ten were played against the computer. In the person counterpart condition subjects were told they would play once against each of ten different people, but in fact the experimenter determined the sequence of offers they would face in order to insure that the person and computer offers were counterbalanced between five 50-50 (fair) offers, one 70-30 (less-fair) offer, two 80-20 (unfair) offers and two 90-10 (unfair) offers.

Behaviourally subjects accepted all of the fair and most of the less-fair offers, but rejected roughly 50 % of the unfair offers by people, while accepting roughly 80 % of the unfair offers by the computer. Sanfey *et al*. find activation in the rostral areas of anterior cingulate cortex and dorsolateral prefrontal cortex, suggesting again the possibility that theory-of-mind may be involved in resolving the conflict between punishing or taking an unfair offer. The strongest evidence that this may be the case is the bilateral activation of the insula, which they interpret as evidence that subjects were experiencing negative emotions (either disgust or anger) with the unfair offers. They found that rejections occurred more often in circumstances where the BOLD contrast value for the right insula activation was greater than the BOLD contrast value for the dorsolateral prefrontal activation. However, when dorsolateral prefrontal processing is interrupted using Transcranial Magnetic Stimulation (TMS), Knoch *et al.* (2006) find that subjects were more willing to accept unfair offers.

The results from the Sanfey *et al.* study resulted in a follow-up behavioural study by Xiao and Houser (2005) of emotional expression in the ultimatum game. The authors find that subjects (player 2s) who can express anger to their counterpart (player 1s) for an unequal offer are significantly more likely to then accept an unequal offer.

One interpretation of these results is that we must first form expectations of what offers to anticipate and accept from our different counterparts. When we get an unfair offer, contrary to our expectations, error-predictor processes are activated through our emotions, as seen by the insula, setting up the possibility for rejection. Insula activation causes a conflict between our simple desire to get as much money as we can or to punish an unfair offer. Emotional expression may bias this conflict in favour of accepting. But what is considered "unfair" is likely to depend on the subjects' understanding of the intentions of player 1. It may be that dorsolateral prefrontal cortex is necessary for the encoding of inequity in the brain.

The investment game

In the double-blind investment game, studied by Berg *et al.* (1995), two persons are given $10 as a show up-fee. Person 1 is then given the opportunity to send none, some, or all of his or her $10 to person 2. Whatever amount of money is sent will triple, e.g. if person 1 sends all $10, then person 2 will get $30. Person 2 then decides how much of the tripled money to send back to person 1.

The sub-game perfect equilibrium prediction is that person 2 should keep all the money, and therefore person 1 should send nothing. Alternatively, social norms may exist that interpret sending money as an obligation for player 2 to reciprocate. All but two of the 32 first movers send some amount of money, with two-thirds sending $5 or more, and about half of these high-trust subjects get more sent back to them than they originally sent before tripling.

In McCabe and Smith (2000) a simplified form of the investment game is introduced. The “voluntary trust game” has only two moves for each player. Player 1 can choose to end the game by moving left, giving each player $10, or by moving down. If player 1 moves down, player 2 can move right, giving player 1 $15 and player 2 $25, or play down, resulting in $0 for player 1 and $40 for player 2. The move down by player 1 is risky, and can be interpreted as trusting player 2, since player 1 gives up $10 and may get $0. Similarly, the move right by player 2 is interpreted as being trustworthy since player 2 gives up $40 and only receives $25. When played as a one-shot game by subjects half of the player 1s are trusting and three-quarters of the player 2s, who then get to move, are trustworthy.

To test whether a theory of mind may be important in playing trust games, McCabe *et al.* (2003) compare behaviour in the standard voluntary trust game to behaviour in an involuntary trust game where player 1 is forced to move down, and player 2 is told this. The authors find that player 2s are twice as likely to make the trustworthy move in the voluntary trust game compared to player 2s in the involuntary trust game. They hypothesize that the increased propensity to move right in the voluntary trust game occurs because player 2 infer player 1′s intentions to cooperate when player 1 has given up a sure thing in order to make them both better off. This inference is what leads to greater trustworthiness.

Kosfeld *et al.* (2005) demonstrate that the brain distinguishes between social trust, and monetary risk-taking, by having subjects inhale intranasally the synthetic neuropeptide oxytocin. They found that subjects who inhaled the oxytocin were more likely to invest more (trust more) in an investment game than subjects who inhaled an inert control. However, subjects showed no differences in their risk preferences over gambles. Oxytocin is known to facilitate social approach in some animals and it seems may play a similar role in trust relationships.

If the investment game is repeated it allows subjects to form a reputation with respect to a behavioural type (such as being a trusting or trustworthy individual). This in turn allows their partners to form beliefs over such types. In a study by King-Casa *et al.* (2005), the authors study subjects' BOLD responses to repeated play of the investment game. The authors find that responses in the caudate nucleus of the second mover was greatest when the first mover invested more in response to the second movers previous reciprocity. Furthermore, the second movers' intention to reciprocate was observed as a shift in peak activity in the caudate nucleus from the time when the second mover saw the first mover's decision to before the first mover's decision suggesting that second movers learn to anticipate first movers' trustworthiness.

In a more recent study, Krueger *et al.* (2007) do an fMRI study of reputation building in the brains of first movers in a repeated voluntary trust game. They find evidence that two different brain systems may be used to develop first mover trust. A personal trust system involves early use of the paracingulate cortex (theory-of-mind) followed by later activation of the septal region of the brain (oxytocin receptor sites) suggesting that repeated experience with another player's cooperation can lead to the evaluation of that player as a "trustworthy" person resulting in an increased production of oxytocin and allowing greater trust. This is consistent with the finding of Kosfeld *et al.* However, a second trust system seems to be more situational and less personal. This system does not use theory of mind early on but does use the reinforcement learning system to build trust. In brains using this system the authors observed theory-of-mind areas activated in the latter stages of play, but not in early play, suggesting that situational trust uses theory of mind to fine-tune expectations over when a counterpart will defect.

Impersonal exchange and the price system

In impersonal exchange markets people must decide whether to buy or sell goods based on prevailing prices. Prices in turn are discovered by trial and error through the aggregation of messages sent to market institutions. Laboratory market experiments which induce supply and demand curves (see Smith 1982), and study subjects bids and asks in auctions of various kinds, find that a number of auctions, such as the Double-Auction, converge quickly to competitive equilibrium thus maximize market efficiency. Gode and Sunder (1993) show that convergence occurs in the Double Auction, in standard laboratory environments even when humans are replaced with Zero-Intelligence traders, i.e. software that bid and ask randomly. While their results demonstrates the robustness of the auction rules, at least for some environments, it does not answer the question as to how the market affects real players decisions. This is made clear in the study of laboratory stock market bubbles where Smith *et al.* (1988) showed that subjects bids and asks (as well as buy and sell) decisions can easily move away from equilibrium price predictions.

Based on the reinforcement learning theories of the brain Lohrenz *et al.* (2007) hypothesized that stock market decisions might be dependent on the brain's learning strategy. In their experiment 54 subjects made decisions as to how much to be in the market (and how much to be in cash) in 20 decision periods in 20 historical markets. They found that the usual temporal difference learning signal between subjects bet and the market outcome was correlated with BOLD responses in the ventral striatum. But another signal, which they call a fictive error signal, was also shown to be correlated with BOLD responses in the dorsal striatum. The fictive error signal is measured after the choice, and when next period's price is observed, as the difference between the return the subject could have received (if they made their best choice against the now observed market price) minus the return on their actual choice (based on the realized price but unobserved at the time of their decision). The fact that the brain is processing two learning signals in order to decide what to do next can help explain why decision makers can get caught in the moment in a "bull" market.

CONCLUSION

Articles such as this one, or critiques such as that made by Gul and Pessendorfer, are unlikely to change the direction or progress in neuroeconomics. When I attended graduate school in economics there was only one lecture on experimental economics in my whole five years.[3] Since then a Nobel Prize in economics was awarded in 2002 recognizing the role of experiments in economics. Neuroeconomics is at a similar stage, and similarly well-directed criticism that improves practice will be appreciated.

To date neuroeconomics has focused mainly on wants, and economic choices, a little on strategic interaction and personal exchange, and very little on markets. If neuroeconomics is to be more valuable to economics this will have to change over time to a greater focus on how our brains interact with our market institutions.

The prohibitive feature to neuroeconomics experiments is the high fixed cost (in intellectual capital) and the high marginal cost (in scanner time) of running an fMRI or single cell firing experiment. It is difficult to

---

[3] That seminar was given by Charles Plott at the University of Pennsylvania in the theory workshop organized by David Cass and Karl Shell. As part of his talk Dr Plott ran a double auction to illustrate market convergence to equilibrium prices where the market was characterized by private decentralized information about values and costs. The number of questions raised by this simple demonstration, and unanswered by eight years of undergraduate and graduate course work is what perked my interest in experimental economics.

say how much these costs will change over time. Even so many labs are active and incremental knowledge will slowly be gained. At the same time, commercial uses of imaging experiments have begun to emerge including neuromarketing, lie detection, and other forms of employee screening. These enterprises are likely to quickly find imaging designs and data analysis that works for their questions without understanding the underlying neuronal mechanisms at work, but in doing so they may uncover many practical questions that economists find interesting.

A question that is often asked, is when is neuroeconomics going to offer a novel prediction? For example, can neuroeconomics write down a game and predict an outcome that standard game theoretic models cannot? This is difficult to do given the generality of game theory. However, it should soon be possible to use neuroeconomics to specify individual types and, based on a model of incomplete information, make novel predictions. Much more exciting will be if neuroeconomics helps us understand new markets that are just evolving such as the virtual markets in SecondLife® and other online worlds.

REFERENCES

Balleine, B. and A. Dickinson. 1998. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37: 407–19.

Berg, J., J. Dickhaut and K. McCabe. 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10: 122–42.

Botvinick, C. S. C. and J. D. Cohen. 2000. Anterior cingulate and prefrontal cortex: who's in control? *Nature Neuroscience* 3: 421–3.

Camerer, C., G. Loewenstein and D. Prelec. 2005. Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature* 2005: XLIII.

Cherry, T., P. Frykblom and J. Shogren. 2002. Hardnose the Dictator. *American Economic Review* 92: 1218–21.

Coltheart, M. 2004. Brain imaging, connectionism, and cognitive neuropsychology. *Cognitive Neuropsychology* 21: 21–5.

Connover, K. and P. Shizgal. 2005. Employing labor-supply theory to measure the reward value of electrical brain stimulation. *Games and Economic Behavior* 52: 283–304.

Forsythe, R., J. Horowitz, N. Eugene Savin and M. Sefton. 1994. Replicability, fairness and pay in experiments with simple bargaining games. *Games and Economic Behavior* 6: 347–69.

Frith, C. D. and U. Frith. 1999. Interacting minds – a biological basis. *Science* 286: 1692–5.

Gazzaniga, M., R. Ivry and G. Mangun 2002. *Cognitive neuroscience*, 2nd Edn. New York: W. W. Norton & Company.

Gigerenzer, G. and R. Selten. 2001. *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.

Glimcher, P. W. and A. Rustichini. 2004. Neuroeconomics: The concilience of brain and decision. *Science* 306: 447–52.

Gode, D. and S. Sunder. 1993. Allocative efficiency of markets with zero intelligence traders: market as a partial substitute for individual rationality. *Journal of Political Economy* 101: 119–37.

Gul, F. and W. Pesendorfer. 2005. The case for mindless economics. Working Paper. Princeton, NJ: Princeton University.

Guth, S., R. Schmittberger and B. Schwarze. 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3: 367–88.

Harley, T. 2004. Does cognitive neuropsychology have a future. *Cognitive Neuropsychology* 21: 3–16.

Henson, R. A. 2005. What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology* 58A: 193–233.

Henson, R. N. A., Rugg, M. D., Shallice, T., Josephs, O. & Dolan, R. 1999. Recollection and familiarity in recognition memory: An event-related fMRI study. *Journal of Neuroscience* 19: 3962–72.

Hoffman, E., K. McCabe and V. Smith. 1996. Social distance and other-regarding behavior. *American Economic Review* 86: 653–60.

Hoffman, E., K. McCabe, K. Shachat and V. Smith. 1994. Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior* 7: 346–80.

Holland, J. H., K. J. Holyoak, R. E. Nisbett, R. E. Thagard. 1986. *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Houser, D. 2008. Experimental neuroeconomics and non-cooperative games. In *Handbook of neuroeconomics*, ed. P. W. Glimcher, C. F. Camerer, E. Fehr and R. A. Poldrack. Academic Press.

Houser, D., D. Schunk and E. Xiao. 2007. Combining brain and behavioral data to improve econometric policy analysis. *Analyse & Kritik* 29: 86–96.

Huetell, S., A. Song and G. McCarthy. 2004. *Functional magnetic resonance imaging*. Sunderland, MA: Sinauer Associates.

Kable, J. and P. Glimcher. 2007. The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience* 10: 1625–33.

King-Casa, B., D. Tomlin, C. Anen, C. Camerer, S. Quartz and P. Read Montague. 2005. Getting to know you: Reputation and trust in a two-person economic exchange. *Science* 308: 78–83.

Knoch, D., A. Pascual-Leone, K. Meyer, V. Treyer and E. Fehr. 2006. Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314: 829–32.

Knutson, B., A. Westdorp, E. Kaiser and D. Hommer. 2000. FMRI visualization of brain activity during a monetary incentive delay task. *NeuroImage* 12: 20–7.

Knutson, B., C. M. Adams, G. W. Fong and D. Hommer. 2001a. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience* 21(RC159): 1–5.

Knutson, B., G. W. Fong, C. M. Adams, J. L. Varner and D. Hommer. 2001b. Dissociation of reward anticipation and outcome with event-related FMRI. *NeuroReport* 12: 3683–7.

Kosfeld, M., M. Heinreichs, P. Zak, U. Fishbacher and E. Fehr. 2005. Oxytocin increases trust in humans. *Nature* 435: 673–6.

Krueger, F., K. McCabe, J. Moll, N. Kriegeskorte, R. Zahn, M. Strenziok, A. Heinecke and J. Grafman. 2007. Neural correlates of trust. *Proceedings of the National Academy of Sciences, USA* 104: 20084–9.

Kuhnen, C. and B. Knutson. 2005. The neural basis of financial risk taking. *Neuron* 47: 763–70.

LeDoux, J. 2002. *Synaptic self*. New York: Penguin Books.

Ledyard, J. 1986. The scope of the hypothesis of Bayesian equilibrium. *Journal of Economic Theory* 39: 59–82.

Logothetis, N., J. Pauls, M. Augath, T. Trinath and A. Oeltermann. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412: 128–30.

Lohrenz, T., K. McCabe, C. Camerer and P. Read Montague. 2007. Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences, USA* 104: 9493–8.

Maynard-Smith, J. 1982. The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology* 47: 209–21.

McCabe, K. 2002. Neuroeconomics. In *Encyclopedia of Cognitive Science*, ed. L. Nadel. London: Nature Publishing Group.

McCabe, K. and V. Smith. 2000. A two person trust game played by naïve and sophisticated subjects. *Proceedings of the National Academy of Sciences, USA* 97: 3777–81.

McCabe, K., M. Rigdon and V. Smith. 2003. Positive reciprocity and intentions in trust games. *Journal of Economic Behavior and Organizations* 52: 267–75.

McCabe, K., D. Houser, L. Ryan, V. Smith and T. Trouard. 2001. A functional imaging study of 'Theory of Mind' in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences, USA* 98: 11832–5.

McClure, S., J. Li, D. Tomlin, K. Cypert, L. Montague and R. Montague. 2004a. Neural correlates of behavioral preference for culturally familiar drinks. *Neuron* 44: 379–87.

McClure, S., D. Laibson, G. Lowenstein and J. Cohen. 2004b. Separate neural systems value immediate and delayed monetary rewards. *Science* 306: 503–7.

Montague, R. 2006. *Why choose this book*? New York: Dutton, Penguin Group.

NIDA, NIMH, NIDDK, sponsors. 2002. Special review issue on reward and decision. *Neuron* 36: 2.

North, D. 2005. *Understanding the process of economic change*. Princeton, NJ: Princeton University Press.

O'Doherty, J., P. Dayan, J. Schultz, R. Deichmann, K. Friston and R. J. Dolan. 2004. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304: 452–4.

Preuschoff, K., P. Bossaerts and S. Quartz. 2006. Heural differentiation of expected reward and risk in human subcortical structures. *Neuron* 51: 381–90.

Sanfey, A., J. Rilling, J. Aronson, L. Nystrom and J. Cohen. 2003. The neural basis of economic decision making in the Ultimatum Game. *Science* 300: 1755–1758.

Schultz, W. 2000. Multiple reward signals in the brain. *Nature Reviews Neuroscience* 1: 189–207.

Schultz, W., P. Dayan and R. Montague. 1997. A neural substrate of prediction and reward. *Science* 275: 1593–9.

Simon, H. A. 1957. *Models of man*. New York: Wiley.

Singer, T. 2008. Brain signatures of social decision making. In *Better Than Conscious? Decision Making, the Human Mind, and Implications For Institutions*, ed. C. Engel and W. Singer. Cambridge, MA: MIT Press.

Smith, V. 1976. Experimental economics induced value theory. *American Economic Review* 66: 274–9.

Smith, V. 1982. Microeconomic systems as an experimental science. *American Economic Review* 72: 923–55.

Smith, V., G. Suchaneck and A. Williams. 1988. Bubbles crashes and endogenous expectations in experimental spot asset markets. *Econometrica* 56: 1119–51.

Smith, V. 2008. *Rationality in economics: Constructivist and ecological forms*. New York: Cambridge University Press.

Spitzer, M., U. Fischbacher, B. Herrnberger, G. Gron and E. Fehr. 2007. The neural signature of social norm compliance. *Neuron* 56: 185–96.

Sutton, R. and A. Barto. 1998. *Reinforcement learning*. Cambridge, MA: MIT Press.

Thorndyke, E. L. 1911. *Animal intelligence: Experimental studies*. New York: Macmillan.

Weibull, J. 1996. *Evolutionary game theory*, 2nd Edn. Cambridge, MA: MIT Press.

White, N. M. 1989. A functional hypothesis concerning the striatal matrix and patches: mediation of S-R memory and reward. *Life Science* 45: 1943–57.

Xiao, E. and D. Houser. 2005. Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences, USA* 102: 7398–401.