THE ROYAL SOCIETY

Published online 26 November 2004

Neuroeconomics

Paul J. Zak

Center for Neuroeconomics Studies, Claremont Graduate University, 150 East Tenth Street, Claremont, CA 91711, USA
(*paul.zak@cgu.edu*)

This paper introduces an emerging transdisciplinary field known as neuroeconomics. Neuroeconomics uses neuroscientific measurement techniques to investigate how decisions are made. First, I present a basic overview of neuroanatomy and explain how brain activity is measured. I then survey findings from the neuroeconomics literature on acquiring rewards and avoiding losses, learning, choice under risk and ambiguity, delay of gratification, the role of emotions in decision-making, strategic decisions and social decisions. I conclude by identifying new directions that neuroeconomics is taking, including applications to public policy and law.

Keywords: reward; brain; trust; emotions; strategy; neuroimaging

1. INTRODUCTION

Neuroeconomics is an emerging transdisciplinary field that uses neuroscientific measurement techniques to identify the neural substrates associated with economic decisions. 'Economics' here should be interpreted in the broadest possible sense as any (human or non-human) decision process that is made by evaluating alternatives. A classic non-human example is 'optimal foraging' where, for example, an ungulate must decide when to expend energy to move from the patch of grass it is currently eating to a different location with an uncertain quantity and quality of grass. A human example would be whether to accept a job as a stock analyst at Goldman Sachs for $100 000 per year but with few future pay increases or advancement versus a job as a stockbroker for a small company starting at $40 000 per year but with the potential for much greater income if successful (and the risk of being fired if not). Both of these examples can be expressed mathematically as constrained optimization problems that generate empirically testable predictions. A prediction in the human example is that a person who is more risk averse (in a precisely defined and agreed upon mathematical sense) is more likely to take the 'safe' $100 000 per year job, whereas someone who is less risk averse will gravitate towards the job as a stockbroker.

Economics is typically defined as the science characterizing the optimal allocation of scarce resources. Note: economics is *not* about money (surprisingly, economics has produced very few deep insights about money!) even though money is a convenient way to determine how much someone cares about something. Fundamentally, economics models individuals valuing rewards and choosing among alternatives. I prefer this definition of economics as it maps economic decisions straightforwardly into the neural substrates that produce these decisions. Specifically, each decision involves (i) obtaining information from the environment regarding possible actions, (ii) valuing those actions, and (iii) choosing between them. Each of these three tasks is, in principle, measurable. Further, this hierarchy of how decisions are made can further be broken down into sub-tasks, including determining one's objective(s), filtering incoming information, accessing memories of related events, using heuristics and identifying constraints on cognitive processing (e.g. energy or time constraints). These, too, are measurable.

Neuroeconomics is a natural extension of bioeconomics (Hirshleifer 1985; Gheslin & Landa 1999; Hirshleifer & Zak 2004). The bioeconomics research programme uses evolutionary biology to build models that predict human behaviour (e.g. Zak 2002; Zak & Park 2002). A second progenitor of neuroeconomics is behavioural economics, a field that uses findings from cognitive psychology to better model human decision-making (Camerer 2003). Whereas bioeconomics has focused primarily on ultimate causes of behaviour and behavioural economics has focused on how our evolved psychologies affect decisions, the neuroeconomics research programme seeks to discover proximate causes of choice behaviour. It is proximate causes that probably provide the most leverage when seeking to affect behaviour through policy. For example, introducing laws that seek to influence individual behaviour can be done more effectively and precisely when the proximate mechanisms producing the behaviour are known.

Because of the focus on decisions, neuroeconomics is not limited to studying humans (and should not be). I date the first paper in neuroeconomics as the 1999 *Nature* article by Michael Platt and Paul Glimcher (discussed later), which used an economic approach to understand how rhesus monkeys choose between two cued rewards. Indeed, neuroeconomics is improving research methods and providing new insights on both sides of the shop, i.e. in 'neuro' and in 'econ'. The first plenary meeting of neuroeconomists, organized by Greg Berns of Emory University, was held in autumn 2003. Out of the 30 researchers attending, roughly one-third had a Ph.D. in neuroscience, one-third had a Ph.D. in economics, and one-third had an M.D. This indicates the broad potential of neuroeconomics across disciplines, including clinical applications.

The economics of choice can be broken down into two primary branches, and research in neuroeconomics has a

One contribution of 16 to a Theme Issue 'Law and the brain'.

similar split. The first is solitary choice. Solitary choices are made with little or no input from others and are non-strategic. The job candidate in my example above already has the job offers and must consider which to choose—this is an individual optimization problem. Such problems are represented mathematically by individuals maximizing a 'utility function' subject to a set of constraints (e.g. an income–expenditure constraint, a time constraint, etc.). A utility function, say $U(c)$, is a mapping from consumption of 'stuff', c, into a measure of subjective happiness, U. 'Stuff' can be anything from commodities to sunsets to leisure time. Using a utility function as a person's ultimate objective is consistent with maximizing genetic fitness (Robson 2001; Zak & Denzau 2001).

Predictions from a solitary choice model are made by finding an *equilibrium*, where the preferred choice produces maximal utility subject to the constraints the person faces and the rules governing the environment of choice. The presumption that human beings have a utility function came from descriptions of the behaviour of gamblers by Daniel Bernoulli in the eighteenth century and is the most foundational notion in economics. Solitary constrained utility maximization predicts behaviour in impersonal exchange (e.g. in markets), generally quite well. This model of decision-making works less well when the decision-maker has incomplete or ambiguous information, or is influenced by others' behaviours (herding) or intangibles other than measurable 'stuff' enter into the utility function. Modifications to the classical utility maximization model for these situations have been proposed, but extensive tests of competing models have not produced an accepted new general theory (Kahneman 2004).

The second branch in the study of choice is strategic choice. Continuing with the example of the job seeker, before obtaining the job offer, he or she probably behaved strategically because the rivalry to get a job offer was with other people. Strategies might include buying a new suit to appear professional and successful, wearing a brightly coloured tie or scarf to be memorable, designing a clever resume to generate attention, finding out who the other interviewees are so as to disparage their skills or education to the interviewer, etc. Decisions with socially strategic elements can be described mathematically using game theory. A game-theoretic model of behaviour requires a description of the people in the game, the information each has or can obtain, the actions available to each player, and the pay-off expected from each strategy. A Nash equilibrium of a game identifies an optimal strategy conditional on everyone else in the game also behaving optimally. Game theory models decisions more complex than isolated utility maximization, and its predictive record is more mixed (Camerer 2003).

In summary, economics is the science of decision-making, decisions that both involve others and those that do not. For this reason, economic models can be applied to a wide range of species and behaviours. Neuroscience, on the other hand, has an exquisite arsenal of measurement modalities, but historically has focused on characterizing a quite limited set of behaviours. Therefore, there is a natural affinity between neuroscience and economics as one has produced and tested many behavioural models without asking what produces the behaviour, whereas the other is able to open the black box that generates behaviours but is searching for interesting behaviours to study.

The expected benefits of neuroeconomics on each side of the shop are high. For economics, neuroeconomic research will lead to the building of models that predict economic and social behaviours better and that are grounded in neurobiology. This will allow economists to answer fundamental questions they are unable to address now such as: why do two individuals faced with the same information and incentives make different choices? Why does the same individual sometimes make choices that are inconsistent? How much is choice behaviour affected by childhood development, if at all? Currently, most answers to economic questions focus on average choices, rather than individual or temporal variation in choices, and model building has a 'what-if' quality where new models are often built without any motivating data. In the application of economic models to policy, most laws seek to circumscribe extreme behaviours, not average behaviours, so an understanding of the interpersonal and intertemporal variation in choices is fundamental to effective public policy.

On the neuroscience side, neuroeconomics provides a host of well-studied and (often) interesting decision tasks that are begging to have their neural 'underpinnings' identified. For example, social cognitive neuroscience is an exciting and important new field (Adolphs 2003), and game-theoretic models of social interactions are an obvious source of tasks to study. Economic models supply the structure of the social interaction as well as (usually) field-tested behavioural predictions, saving researchers from having to reinvent the wheel. Such game-theoretic models are often fairly complex, and neuroeconomics is moving neuroscientists to study tasks that approach those that humans actually do in their daily lives. Finally, because economic models have objective behavioural measures, usually involving monetary transfers, neuroeconomic experiments engage subjects' attention better and have added control compared with tasks that are simply passive (e.g. viewing photographs) or in which the subjects are asked to 'imagine' themselves doing something. Most neuroeconomists also follow the ethic in experimental economics that prohibits the deception of subjects. With a guarantee of no deception, subjects make choices without trying to 'game' the experimenters by figuring out what they are 'really' looking for.

2. BASIC BRAIN FACTS AND TERMINOLOGY

There are roughly 100 billion neurons in the human brain, with each neuron directly connected to between 1000 and 10 000 other neurons. Brain tissue can be separated into grey matter (neurons) and white matter (axons and dendrites, the connections between neurons). Grey matter makes up 40% of the brain, but consumes 94% of the brain's oxygen owing to the firing of action potentials (electrical pulses) that allow one neuron to communicate with other neurons. The cortex (from the Latin for bark) is the outer surface of the brain that is used for information processing and higher mental functions. Because the human brain is folded (to pack more cortical tissue into the skull), a brain region may be identified as being on a gyrus (hill, pleural gyri) or in a sulcus (valley, pleural sulci).

The brain is grossly divided into four sections: the frontal, temporal, parietal and occipital lobes (see figure 1). Each lobe performs several functions, containing smaller structures that do specific tasks, often in concert with other brain regions through connections called projections. The brain sits on top of the brain stem, which leads to the spinal column. A cauliflower-shaped structure, the cerebellum, sits below the occipital lobe and adjacent to the brainstem. A common way to identify cortical regions in the brain is by using 'Brodmann's Areas', which are numbered from 1 to 47. These are abbreviated BA*x*, where *x* is the integer corresponding to that region. German physician and anatomist Korbinian Brodmann (1868–1918) identified brain regions based on similar cellular and laminar structures (see figure 2).

Because the brain is three-dimensional, identifying locations requires specialized terminology. Terms for locations of brain regions include: dorsal (top, from the Latin for back); ventral (bottom facing the central axis, from the Latin for belly) or basal; rostral (front, from the Latin for beak) or anterior; caudal (back, from the Latin for tail) or posterior; superior (towards the top); inferior (towards the bottom); medial/mesial (middle); lateral (away from midline); and orbital (above the eyes, from the Latin orbita meaning eye sockets). Generally, brain regions that are ventral and inferior tend to be phylogenetically older than dorsal and rostral regions, with older regions mostly conserved in lower animals.

Much of the nervous system is outside of volitional control (is autonomic). There are two opposing sides to the autonomic nervous system. Sympathetic responses are associated with the four Fs (fright, flight, fight and fornication), whereas the parasympathetic nervous system activates when it is time to rest and digest. The sympathetic is arousing, and the parasympathetic relaxes; maintaining the balance between these sides of the autonomic nervous system is essential to health and growth. The hypothalamus, a basal midbrain structure, exerts primary control over the autonomic nervous system. Most emotional responses are also automatic and rapid. Primary emotional responses emanate from the brain's limbic structures. The limbic system (limbus, Latin for edge) is grey matter in the medial temporal lobe, and includes the amygdala (associated with positive and negative emotions), hippocampus (associated with long-term memory), cingulate cortex (attention and error detection) and olfactory cortex (smell).

(a) *Measurement of brain activity*

Neuroscientists use a variety of measurement modalities to gauge neural activity, including PET, fMRI, EEG/ERP, intra- or extracellular recording of electrical activity of single neurons, bioassays of blood, urine and cerebral spinal fluid, responses to drug infusions, as well as studies of patients with specific central nervous system lesions. Most of the neuroeconomics research performed on humans has used fMRI or PET, both of which provide high spatial resolution of regional brain activity during particular tasks with moderate to low temporal resolution (between 100 ms and 2 s for fMRI, and 30 s or more for PET; see Buckner 2003).

PET imaging was first performed on humans in the early 1970s. Experimental subjects are injected with a radioactive isotope that emits positrons (positively charged electrons). Subjects then lie in a ring of crystal detectors and a camera that captures radioactive decay (when a positron meets an electron they annihilate each other and emit gamma rays). When neurons fire they deplete glucose and oxygen and require increased blood flow to resupply these substances. Blood flows to neurons roughly proportionally to their firing rates. PET measures the accumulation of the radioactive tracer in brain regions; regions metabolizing glucose faster receive more blood flow and emit more gamma rays. A computer algorithm constructs the measurements of regional cerebral blood flow in three dimensions as an indirect measure of neural activity. The use of radioisotopes with short half-lives places a 1 h time limit on PET experiments and restricts subjects to two studies per year.

fMRI was first used on humans in 1992, and produces 3D renderings of regional neural activity. The data obtained by fMRI are BOLD signals that indirectly measure regional neural and synaptic activity by examining the amount of oxygenated to deoxygenated blood (the haemodynamic response). Neural firing increases the demand for oxygenated blood (oxyhaemoglobin). Because deoxyhaemoglobin is paramagnetic, it produces a measurably larger signal relative to oxyhaemoglobin when perturbed by a short radio-frequency pulse. These differences are small and can be measured only in a very powerful magnet (currently MRI scanners used for humans have magnets from 1 to 8 T; a 1 T magnet is 20 000 times stronger than the magnetic field on the Earth's surface). Higher field-strength magnets increase resolution (up to 1 mm^3) but also increase the noise associated with signal detection. This makes the analysis more difficult because external confounds must be eliminated. Magnetic fields are not associated with any adverse health effects (Kangarlu *et al.* 1999), though very powerful magnets (4 T or more) can induce temporary dizziness and a metallic taste in the mouth. fMRI experiments are limited in time only by the subject's ability not to fidget or fall asleep, and can be repeated on the same subject indefinitely.

Both fMRI and PET use a 'subtraction' method to statistically identify regional neural activation during a task. This is done by measuring brain activity during the task of interest and then removing activation measured in a control task. The control task is often 'baseline' neural activity (e.g. staring at a fixation point), though better studies use control tasks that are closer to the task of interest. For example, if the task is to choose between two alternatives involving monetary rewards, a good control task would be giving the subject a monetary reward absent choice. The subtraction then removes the activation in the brain from simply receiving (or anticipating) reward and identifies brain regions active in making the choice. Choosing a good control task is a major feature of these experiments, and readers of this literature should be sceptical of the results if the experimental design is poor. Both PET and fMRI correlate tasks with regional brain activity; demonstrating causation requires others methods discussed below.

Montague *et al.* (2002) at Baylor College of Medicine's Human Neuroimaging Laboratory have provided an important advance to study regional brain activity during social interactions that they call 'hyperscanning.' Hyperscanning allows two or more subjects in MRI scanners in different locations to interact simultaneously through the

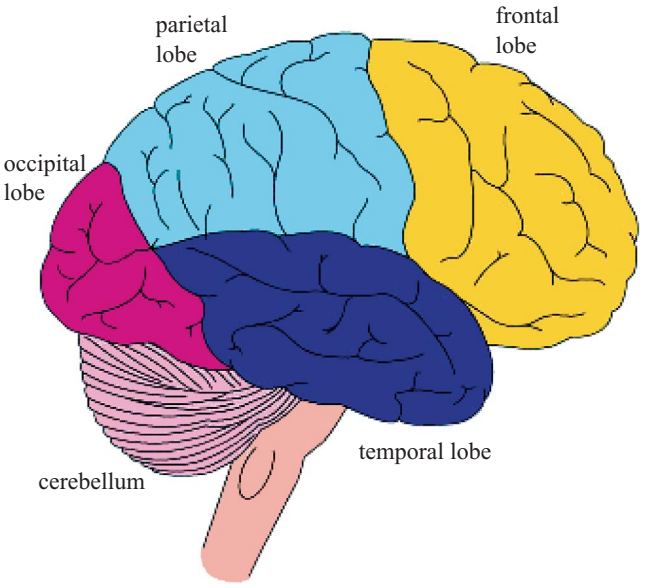

Figure 1. The lobes of the brain, cerebellum and brain stem. (Copyright Mark Dubin; printed here with permission. First published online: See http://spot.colorado.edu/~dubin/talks/brodmann/brodmann.html.)

Internet with behavioural and even visual and auditory feedback between subjects while measuring brain activity. This literally allows researchers to see one person's brain affect another person's brain. So far, Montague and collaborators have hyperscanned eight subjects simultaneously. Their proprietary software synchronizes stimulus presentation and BOLD signal acquisition across subjects and locations. Hyperscanning opens up fMRI from single- to multiple-subject studies and will see increasing use in the coming years to answer questions in social cognitive neuroscience and the neuroeconomics of social decisions.

EEGs/ERPs use between 16 and 256 scalp electrodes to measure the electrical activity of large groups (more than one million) of neurons. EEGs are used clinically to help diagnose neurological disorders, especially epilepsy, by examining the synchronicity, frequency and amplitude of EEG tracings called 'waves' while a patient sits or lies down. ERPs differ from EEGs in that experimental subjects are given specific tasks to do that may provoke regional brain activation. The characteristics of ERP waves identify regional excitatory or inhibitory neural activity. ERPs provide higher temporal resolution than fMRI or PET (*ca.* 10 ms) but lower spatial resolution. The other advantages of ERP over fMRI or PET are its relatively low cost, less demanding statistical analyses (two dimensional versus three dimensional), and greater freedom of movement for subjects. The disadvantages of ERPs include low spatial specificity, subject performance fatigue that occurs because many trials are required per subject to reduce background noise and artefact, and potential problems with inter-subject comparisons because consistent electrode placement depends on a careful identification of bony landmarks that vary across subjects. Some laboratories have now combined ERP and fMRI to obtain high temporal resolution together with high spatial resolution.

Measuring the firing rates of single neurons in the brain requires that a microelectrode be attached to, or inserted into, the neuron cell body. Neuron cell bodies vary in size from 4 to 100 μm (a micrometre is one thousandth of a millimetre), and obtaining internal or external recordings from a neuron often damage or destroy it. Single neuron firing measurements offer the highest level of spatial specificity, but are seldom performed on humans. Some surgical patients have electrode grids place on the convexity of the brain or deeper inside the brain ('depth electrodes') that measure the activity of a few neurons, and these patients have occasionally been used in research. Animals are more commonly used when recording the firing of single neurons.

Bioassays provide an indirect measure of neural activity, and have the advantage of being able to identify cascades of activity that produce behaviour, as well as facilitating the investigation of individual-specific confounds. Obtaining biological material, such as blood, is invasive and the act of obtaining the sample may affect what is being measured (e.g. hormones or neurotransmitters). Combining bioassays with other measurement techniques allows researchers to triangulate neural activity within a single experiment.

Using pharmaceuticals in experiments is an important method to induce behaviour, i.e. to move from correlation to causation, and its use in neuroeconomics is just beginning. Similarly, comparing the behaviour of patients with focal brain lesions with healthy controls is also an important step in establishing the necessity of a brain region for a particular behaviour. Several laboratories, including my own, are studying brain-damaged patients but have not yet published their findings. Temporary brain lesions or neural hyperactivation can be induced by focusing a magnet field on the convexity of the brain using TMS. I am not aware of any neuroeconomics experiments using TMS, but it is an important (though not completely risk-free) technique that can be used to ascertain causation.

3. MAJOR FINDINGS IN NEUROECONOMICS

The research topics studied by neuroeconomists fall into two major categories: (i) identifying the neural processes involved in decisions in which standard economic models predict behaviour well; and (ii) studies of 'anomalies' where the standard models fail. For the latter, often several alternative models have been proposed with different behavioural assumptions that predict decisions equally well and therefore the 'true' sources of behaviour are unknown (Camerer 2003). Research in category (i) is often headed by a neuroscientist or an M.D., where much of the research in (ii) is led by economists. Many research teams now include both economists and neuroscientists/M.D.s and consequently the breakdown of research into these two categories is beginning to blur. Because of the rapid growth of the neuroeconomics literature, the review here will be incomplete by the time this issue goes to press, but I maintain an updated neuroeconomics reading list at my laboratory Web site, http://www.pauljzak.com.

(a) *Reward acquisition*

All animals need to obtain resources to survive, and the neural structures needed for reward acquisition are primitive and well conserved across species. Choice execution is preceded by the evaluation of the reward associated with each choice, but the evaluative substrate is unknown. Platt & Glimcher (1999) trained rhesus monkeys in a colour-

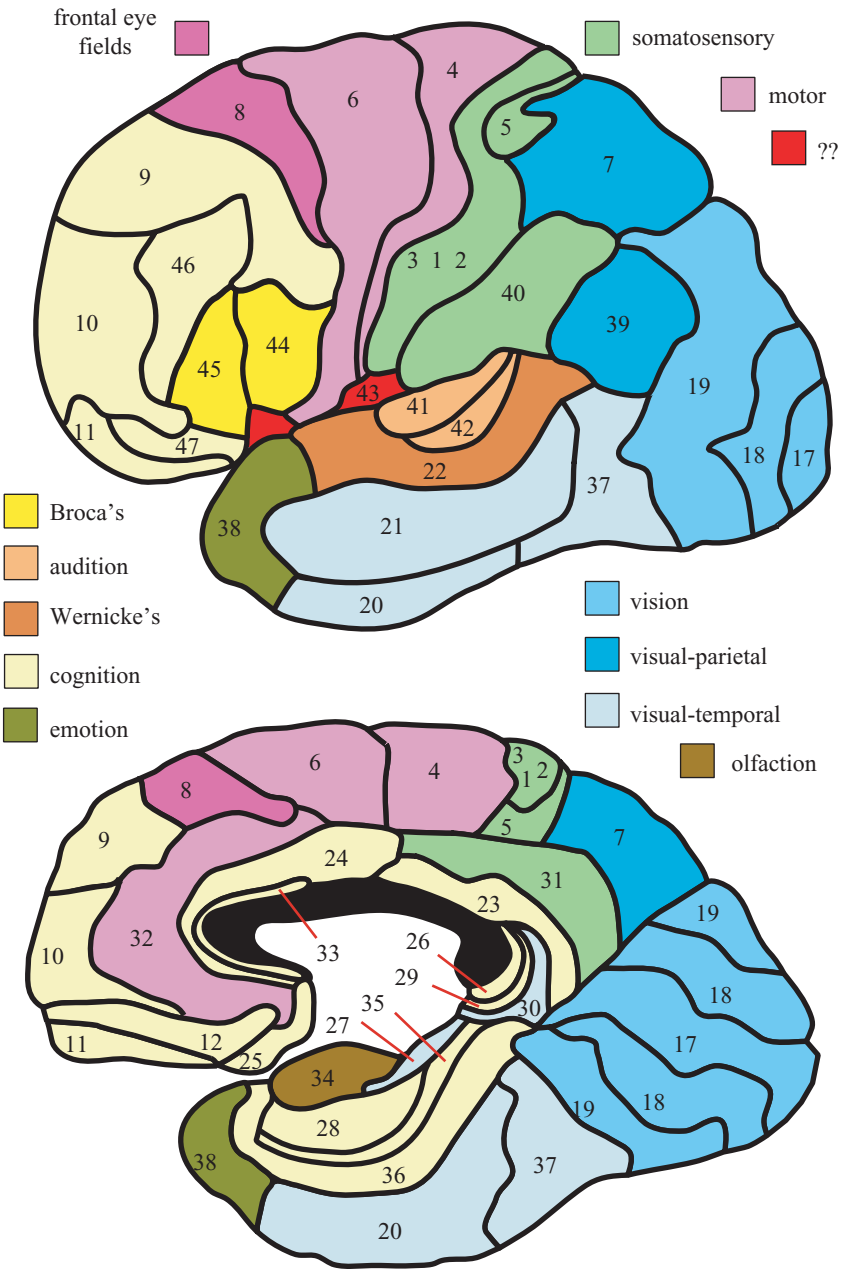

Figure 2. Brodmann's areas. (Copyright Mark Dubin; printed here with permission. First published online: See http://spot.colorado.edu/~dubin/talks/brodmann/brodmann.html.)

cued eye saccade task. The correct left or right saccade was rewarded with a squirt of juice. These researchers suspected that area LIP was being used to evaluate rewards as projections from the visual cortex converge in area LIP before being relayed to the motor cortex for execution. Platt and Glimcher measured the firing rate of 40 neurons in area LIP in three monkeys as they varied the juice reward for the correct saccade either in absolute amount, or probabilistically (i.e. for the latter, each correct saccade was rewarded with juice with a given probability). They found that 62.5% of area LIP neuron activation was correlated with expected gain. These findings for area LIP were recently replicated and extended by Newsome's laboratory (Sugrue *et al.* 2004).

Glimcher *et al.* (2004) go further, to argue that the utility function that economists presumed existed to explain behavioural data is a physiological reality in area LIP. That is, area LIP neurons do not behave 'similar to' a utility function, but 'are' a physiological utility function in monkey brains (i.e. area LIP neurons perform the calculations needed to determine utility). However, this does not preclude the existence of other brain regions that are utility functions (see below). Glimcher *et al.* (2004) support this claim by showing that area LIP firing rates can be used to predict the behaviour of monkeys in several reward acquisition tasks. Work with humans using fMRI is currently underway in Glimcher's laboratory to determine if the human homologue of area LIP is also a physiological utility function (Nelson *et al.* 2004).

Reward acquisition requires a motivating mechanism to obtain the reward as well as the ability to predict reward size to gauge the effort needed to pursue the reward. Schultz *et al.* (1997) review single-neuron firing studies of juice rewards in non-human primates and identify dopami-

nergic neurons in the ventral tegmental area and substantia nigra as processing rewarding stimuli, activating during novel stimuli, and most importantly, firing proportional to the error of the actual to the expected reward. They introduce the temporal difference mathematical model to show how dopamine neuron activity can be used to predict an animal's behaviour as it learns about rewards.

Dopaminergic neurons are particularly dense in the nucleus accumbens in the ventral medial region, and this region has strong projections to the medial forebrain, which is active in many decision-making tasks. Although cocaine, methamphetamines, humour, and even viewing faces of attractive women by heterosexual men produce acute activation in the nucleus accumbens (Aharon *et al.* 2001; Mobbs *et al.* 2003), recent experiments have shown that dopamine release is not the same as pleasure (Garris *et al.* 1999). Indeed, activation in the nucleus accumbens and ACC (BA23/24/31/31/32) is associated with attentional demands. Breiter *et al.* (2001) used event-related fMRI to examine regional activation to the expectation and realization of monetary gains and losses for 12 human subjects. Monetary awards were made without any subject choice in this experiment. They showed that expected and actual rewards were associated with significant haemodynamic responses in the SLEA and orbital gyrus. In addition, activation in the nucleus accumbens, SLEA and hypothalamus tracked the highest monetary values. Gains produced predominant activation in the right hemisphere (particularly the nucleus accumbens and hypothalamus), whereas losses produce greater left hemisphere activity (especially the left amygdala). These findings appear to indicate that gains produced neural rewards, whereas losses provoked emotional responses associated with fear or regret.

Knutson *et al.* (2001) further dissociate the anticipation of reward with its realization by having nine subjects respond with a button push to a coloured cue in an fMRI study. A rapid button push for a yellow cue produced a $1 reward, a rapid response to a blue cue was not rewarded, and a red cue required no response. After each trial, subjects were told how much they earned on that trial and in total. Knutson and colleagues acquired fMRI signals, before and after subjects received feedback on reward or no reward. Anticipation of reward produced activity in the dopamine-receptor-rich ventral striatum (consisting of the substructures caudate nucleus and putamen), whereas notification that a reward was earned (approximating reward consumption) produced primary activation the MPFC.

In a follow-up study with a larger reward ($5), Knutson *et al.* (2003) show that the MPFC (BA 10/32), posterior cingulate cortex (BA 26/30) and parietal cortex (BA 7) activate during the notification of a monetary reward. Interestingly, when rewards were anticipated but *not* obtained, the MPFC showed decreased activation relative to baseline (no outcome). The MPFC has the densest dopaminergic innervation of any cortical region and Knutson and colleagues argue that this region serves as a utility function, whereas the nucleus accumbens guides reward anticipation and learning. An excellent review of this literature is Knutson & Peterson (2004) where the authors make the point that subjective states associated with utility must have an emotional basis—utility must be felt to be valuable—and the MPFC and the OFS circuit appear to map 'wanting' into 'having'.

Montague & Berns (2002) also review the reward and prediction literature. They propose a predictor-valuation model for reward that uses the OFS circuit. Similar to Glimcher's claim for area LIP and Knutson's promotion of MPFC, Montague and Berns provide an array of evidence that OFS values rewards (and punishments). They also provide evidence that reward/punishment evaluation in OFC is separate from the error prediction feature of midbrain dopamine neurons that innervate it.

Dickhaut *et al.* (2003) had nine subjects choose between pairs of lotteries in a PET study. Some of the lotteries produced gains whereas others produced losses (subjects received an initial endowment of $190). Behaviourally, they find risk aversion over gains but not losses, with average response times for loss lotteries 500 ms slower than choices over gains. When compared with a risky reference lottery, gains minus losses produced OFC activation. By contrast, when the reference lottery was a certain payment, gains minus losses produced primary activation in the cerebellum and parietal cortex. Losses minus gains activated dorsal parietal and frontal cortices whether the reference lottery was risky or certain. This report demonstrates how varying the stimulus and/or measurement modality can produce quite different regional activation maps than other similar studies have found. Interpretive caution is called for.

All reward evaluation requires 'emotion' in that ventromedial areas associated with dopamine activate to motivate subjects to acquire resources, and dopamine-innervated cortical regions appear to value resources. It is possible that OFS, MPFC and area LIP *all* value rewards (i.e. are physiological utility functions), with an undiscovered brain region (perhaps prefrontal) determining final valuation when these regions provide conflicting assessments. The asymmetry between gains and losses is also an issue requiring further study by, for example, replicating some of the experiments discussed in this section. Finally, additional research is needed to elucidate the temporal dependence of subcortical and cortical circuits identified in reward evaluation and consumption.

(b) *Certainty, ambiguity and gratification delay*

Neuroscience research has shown that emotions are an important physiological guidance system for choice. For example, Damasio (1994) reported the inability of patients with selective damage to the OFC to execute choices. Kahn *et al.* (2002) showed that amygdala activation was predictive of an anticipated loss. Emotional activation during decisions may be more likely to occur with incomplete information, risk, or choice in a social context. For tasks in which the best decision is difficult to determine through cogitation, emotional markers provide additional information that can guide choice.

The suppression of limbic responses may be part of what makes human choice different from choice by animals. This was investigated in a fascinating field study by Lo & Repin (2002). These researchers proposed that professional foreign exchange traders would have emotional responses to market volatility while trading. With permission from a Boston brokerage firm, they 'wired up' 10 traders for 1 h each to obtain data on six physiological

measures while the traders managed currency contracts of one million US dollars and larger. Lo and Repin simultaneously measured activity in the currency markets. All traders exhibited heightened cardiovascular and electrodermal states during periods of market volatility. More generally, rapid market movements provoked traders' sympathetic nervous systems; this can be interpreted as emotional responses. Interestingly, longer job tenure was associated with reduced sympathetic responses for a given amount of market volatility. This suggests that either experienced traders learned over time to suppress their emotions, or that more emotionally reactive traders left to take other less personally stressful jobs. Lo and Repin were not allowed to obtain data on traders' performance in markets, so we do not know if emotional responses diminished (or improved) the ability to make money. These researchers are currently examining this issue by bringing professional money managers into the laboratory and requiring them to trade to earn monetary returns in simulated markets.

Smith *et al.* (2002) examined the same data as Dickhaut *et al.* (2003) but investigated the role of ambiguity. An ambiguous lottery is one in which the likelihood of one or more of the pay-offs occurring is not fully specified. For example, the subject is asked to choose between lotteries A and B, where A guarantees a payment of $10, and B pays $20 if a red ball is pulled from an urn, and $0 if a blue ball is pulled; the urn contains 90 balls, and at least 50 are red. (Try this yourself: do you prefer lottery A or B? Most people are ambiguity-averse and choose A.) Smith and colleagues report strong activation in the OFC and intraparietal sulcus for gains subtracted from losses without ambiguity. Subtracting risky losses from gains after removing ambiguous lottery choices produced activation in the cerebellum and dorsomedial cortex. This suggests that losses activated cortical regions associated with calculation, while gains activated the older ventromedial system. Ambiguity alone produces small amounts of ventromedial and limbic activation.

Unpublished research by Rustichini *et al*. (2004) used a similar paradigm with 12 subjects choosing between 96 pairs of certain, risky, ambiguous and partly ambiguous lotteries in a PET study. Subjects showed strong ambiguity aversion, but ambiguous and partly ambiguous choices did *not* generate activation in brain regions associated with emotions (e.g. OFC or amygdala). Rather, ambiguous choices were associated with rostrofrontal activation, with substantial deactivation in ventromedial regions. Similar to work from Glimcher's laboratory, Rustichini *et al.* (2004) find strong parietal activation when subjects chose the certain lottery (but they did not explore a parametric relationship between activation and reward amount). There is no consistency between the findings of Rustichini *et al.* (2004) and Smith *et al.* (2002) about the neural substrates associated with ambiguity during choice. I consider this issue important and unresolved.

A major behavioural difference separating humans from other animals is our ability to postpone current gratification for a later (larger) reward. Behaviourally, humans exhibit a strong desire for current reward and rapidly devalue future rewards (Laibson *et al.* 1998). Recent work by McClure *et al.* (2004) used fMRI to examine how the brain decides between current versus delayed rewards. In this study, all rewards were monetary, with current rewards paid immediately after scanning, and delayed rewards paid between two and six weeks later. Delayed rewards always exceeded current rewards. McClure and collaborators found that immediate reward primarily activated the ventral striatum, medial OFC and medial prefrontal cortex. Delayed rewards differentially activated the lateral prefrontal cortex and inferior parietal cortex. These areas were particularly active when the difference between immediate and postponed rewards was small. The authors conclude that choosing between immediate and delayed gratification constitutes a battle between limbic structures that activate for current reward and newer cortical regions that evaluate trade-offs.

(c) *Learning and strategy*

Both the dopaminergic system and emotional responses are important in learning what is valuable or dangerous as animals navigate the world. These systems, and others, update memories of past experiences using the present experience so the animal has a basis for making informed future decisions. In a very careful study, Barraclough *et al.* (2004) investigated reinforcement learning and reward encoding in two rhesus monkeys trained to play a variant of 'matching pennies' against a computer using three different strategies. Matching pennies is a very simple game in which optimal behaviour is a 'mixed strategy' or randomization over choices. The canonical game has two opponents choosing to show either a head or a tail on a penny, and putting the coins down simultaneously on a table. If both pennies show the same face (i.e. either both heads or both tails), player A wins the pennies; otherwise player B wins. The monkeys did this task using eye saccades and juice rewards.

Barraclough and colleagues found that for all the algorithms they used, monkeys learn very quickly to behave optimally by randomizing their choices. A reinforcement learning statistical model fitted the monkeys' choices quite well showing that the history of play by the computer affected the monkeys' current choices. These researchers also recorded the firing of 132 separate neurons in the DLPFC during monkey choices. The firing rate of 37% of DLPFC neurons measured was affected by the previous reward, while the firing rate of 39% of these neurons was influenced by the previous choice. This indicates that the DLPFC may be part of the neurophysiology of reward acquisition, especially when this involves memory-dependent strategic decisions. In humans, the DLPFC, which activates during working memory tasks, may be another physiological utility function. That is, the current value of a reward may be affected by the memories of obtaining similar rewards. If this result is confirmed by other studies (especially in humans), it suggests an important modification to the classical economic model of utility.

Learning involves, of course, more than one brain area and more than one neurotransmitter. For example, the neurotransmitter glutamate and *N*-methyl-D-aspartate receptors are critical for the neural basis of learning in which connections between neurons are strengthened, called LTP (see Riedel *et al.* 2003). Reinforcement learning also appears to require neural activation in the amygdala and OFC (see the excellent review and a proposed mathematical model in Dayan & Balleine (2002)). Future

neuroeconomic research on learning should explore the roles of glutamate and LTP.

(**d**) *Cooperation*

Intraspecies cooperation with non-kin is an issue that has attracted substantial attention but is still not well-understood (Boyd *et al.* 2003; Brosnan & de Waal 2003). Particularly stark is costly cooperation in one-shot interactions with the opportunity to defect without punishment. Even in this setting, humans are highly cooperative (Smith 1998; Fehr & Rockenbach 2003). The ability to cooperate has, potentially, positive and negative neural reinforcers. The positive is the (internal and external) reward obtained by being cooperative. The negative may be the neural correlates associated with the loss of a larger reward and the neural activity resulting from social condemnation by one's trading partner after being uncooperative (for a mathematical model of prosocial emotions see Bowles & Gintis (2003)).

Neuroeconomists have sought to identify the neural substrates associated with cooperative behaviour. An early and important contribution by McCabe *et al.* (2001) reported fMRI data for subjects interacting in real-time by computer with another person outside the scanner. McCabe *et al.* (2001) hypothesized that cooperative behaviour would require that subjects use a brain region associated with 'theory of mind' in which a person is able to anticipate what another will do by imagining himself/herself in the same situation. Most humans, except those under 5 years old and most autistics, have a fully operational theory of mind, and it has been localized to include a region in the medial OFC (BA10) as well as several other regions (Frith & Frith 2003). McCabe *et al.* (2001) provided an incentive for cooperative behaviour by using a binary choice version of the 'trust game' (Berg *et al.* 1995) where subjects can earn more money if they cooperate, but cannot communicate except by transferring money to each other through their choices. Subjects denoted DM1 and DM2 made sequential choices for the dollar amounts shown in figure 3, alternating the roles of DM1 and DM2. In figure 3, DM1 either ends the interaction by providing pay-offs of \$0.45 for DM1 and DM2 (moving left), or transfers control to DM2 (moving right). When DM1 yields control of the game to DM2, he or she signals trust in DM2. DM2 then can be trustworthy, earning \$1.80 for DM1 and \$2.25 for DM2 (left), or can be non-trustworthy causing DM1 to earn \$0, and DM2 to earn \$4.05 (right). Note that the 'pie' increases from \$0.90 to \$4.05 (450%) when DM1 chooses to transfer control to DM2.

In a conjunction analysis of cooperative moves by DM1s and DM2s, McCabe *et al.* (2001) find that BA10 is indeed more active (i) than when subjects were not cooperative, and (ii) relative to a control task where subjects were informed that they were interacting with a computer that moved left or right with known probabilities. The authors argue that BA10 is part of the neural architecture that allows gratification delay in order to obtain larger rewards through cooperation. A possible confound in this study is that to generate sufficient fMRI signal, DM1–DM2 pairs made 80 choices in the same dyad so subjects were able to build reputations for cooperation during the experiment. It is also worth mentioning three important aspects of this study. First, there was no deception: the DM in the MRI scanner actually interacted in real time with another human being (a reasonably difficult technical hurdle). Second, the control task was identical to the treatment task but simply removed the intentionality associated with decisions. This allowed these researchers to cleanly extract the neural components of intention. Third, the neuroanatomical hypothesis for activation in BA10 allowed the acquisition of fMRI data optimized for high signal : noise in the region of interest providing higher-quality data.

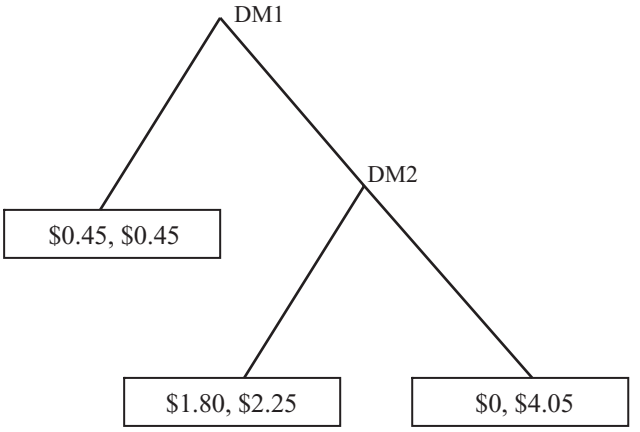

Figure 3. The binary-choice trust game. Dollar figures are, respectively, pay-offs for DM1 and DM2 at each node.

The binary trust game is an iterated PD, where DM1 and DM2 choose to either cooperate or defect. A PD is a strategic interaction in which both parties gain by behaving cooperatively, but are unable to coordinate cooperation; the dominant strategy (choice) is for both DMs to choose to be non-cooperative ('defect'), injuring both DMs by producing low or negative pay-offs. Rilling *et al.* (2002) examined cooperation versus defection when 36 female subjects played 20 or more rounds of the trust game against a human or computer opponent programmed to react in several ways to the other's choices (e.g. tit-for-tat). Removing the pure monetary effect using a condition where the subject knows she is playing against a computer (for the same dollar amount), the social aspect of cooperation produced activation in the anteroventral striatum, right ACC and OFC.

The conclusion from this study is that cooperation is rewarding (striatum), requires attention and the mediation of the conflicting concerns of making more money but behaving in socially less acceptable ways (ACC), and has an emotional component (OFC). Defection by DM1 with cooperation by DM2 was associated with deactivation of the striatum, with a similar deactivation when choosing the cooperative node with a computer partner. The region with the strongest activation during cooperation is the somatosensory association cortex (BA7), consistent with Antonio Damasio's somatic marker theory (Damasio 1994) linking emotions 'experienced' in the body with decisions. (The somatosensory association cortex in the posterior parietal lobe activates during memory, attention and emotional responses to objects.) Rilling *et al.* (2002) partly replicate the finding of McCabe *et al.* (2001) of BA10 activation, but only when subjects cooperated while playing a computer that also moved to the cooperative node, but not in the human–human treatment.

Sanfey *et al.* (2003) used fMRI to analyse another economic task involving cooperation, with their major results associated with the consequences of not cooperating. Sanfey and colleagues had subjects make decisions in the ultimatum game, a sequential decision task to determine the split of a sum of money between two people. For example, DM1 is given $10 and told to offer an integer split to DM2, without seeing or communicating with him or her. DM2 can then accept the split and the amounts are paid, or can reject the split and both DMs earn nothing. Behaviourally, when DM1s offer less than 30% of the money, DM2s nearly always reject offers. From a purely economic point of view, a rejection of any money is 'irrational' because some money is expected to be preferred to none, but humans are social beings and there is clearly a social aspect to this game. DM2s also report feeling angry when a DM1 offers a stingy split.

Sanfey *et al.* (2003) modified the ultimatum game to generate only the following DM1–DM2 offers: {$5, $5}, {$7, $3}, {$8, $2} and {$9, $1}. Only DM2s were scanned. A computer played the role of DM1, but Sanfey *et al.* (2003) deceived subjects into believing that they were playing with another human to simplify the protocol (all subjects reported that they believed this). On some trials, the researchers told DM2s that they were playing against a computer as a control task. Unfair offers differentially activated the anterior insula, DLPFC and ACC. Activation in all three regions was greater for unfair offers from humans than from a computer. Their major finding is that insula activation increased with the unfairness of the offer from a human. Insular cortex activation has previously been associated with disgust, pain, hunger and thirst. Sanfey and colleagues concluded that low offers in the ultimatum game are rejected because of a sense of disgust, while DLPFC activation may be signalling the importance of acquiring money.

Interpersonal trust is the most powerful predictor at the country level of whether nations will experience rising living standards or will remain trapped in poverty (Zak & Knack 2001). Zak *et al.* (2004) examine whether there is a physiological correlate associated with the receipt of a signal of trust that motivates individuals to be trustworthy (that is, to reciprocate trust). Drawing on research with rodents on social recognition and attachment, Zak and colleagues proposed that the neuroactive hormone OT would process signals of trust and induce trustworthy behaviour. They used a variant of the trust game in which all DMs received a $10 show-up payment and were randomly assigned to dyads. DM1s were prompted to send an integer amount, including zero, of their $10 show-up money to the DM2 in their dyad. The amount sent was removed from DM1's account, and was tripled in DM2's account. DM2s were then told the tripled amount that they were sent and the total in their accounts. Next, DM2s were prompted to send some amount back to the DM1 in their dyad, including zero. All interactions were mediated by computer, and subjects were fully informed of the structure of the interaction and the consequences of their choices. Participants where also told that they would only make a single decision.

In this experiment, DMs made decisions serially, and immediately after each decision went to an anteroom and had 28 ml of blood drawn from an antecubital vein. All experiments began at 13.00, a trough in diurnal hormone variation. Zak and colleagues showed that DM2s receiving trust signals had OT levels almost twice that of DM2s in a control task in which DM2s received random (unintentional) monetary transfers of the same average amount as in the treatment task. In addition, higher OT levels in DM2s were strongly associated with trustworthy behaviour. None of nine other hormones measured, except for progesterone, responded to the trust signal nor were associated with DM2 behaviour. Women in the study who were ovulating (progesterone level more than 3 ng ml^{-1}) were less trustworthy than other subjects. Progesterone is known to inhibit OT uptake. This finding indicates that OT is the primary hormone responding to signals of trust (i.e. the behavioural effect is caused by OT and not another hormone). There were no overall gender differences. Their analysis shows that OT is released in response to a signal of trust (the experimental state), rather than being a primary trait of subjects (i.e. DM1s with high OT levels did not behave any differently than other DM1s as these subjects did not receive a trust signal). Zak's team concludes that OT, which activates the parasympathetic system and facilitates dopamine release, is a positive physiological motivator of cooperation.

4. THE FUTURE: CONVERGENT EVIDENCE

One of the important lessons neuroscience can teach economics is the necessity of convergent evidence before a finding is accepted as 'proved'. This typically means using different measurement modalities, subject groups (especially atypical groups), and moving from correlation to causation. An example of this research using economic decision tasks but absent neurophysiological measurement is the study of autistics by Hill & Sally (2003). They compared the behaviour of healthy children and adults with age-matched patients diagnosed with autistic spectrum disorder as they made choices in the PD, ultimatum and dictator games. (In the dictator game, DM1 is given a monetary endowment and chooses to give some amount of it to an unknown DM2; DM2 does not make a choice. Healthy adult DM1s typically offer 10% or less to DM2s in this game which is designed to measure altruistic behaviour.) They report that autistics were no less likely to cooperate, but did not learn to be strategic in repeat play as did healthy subjects. Some of this failure to learn strategy appeared to derive from a lack of a theory of mind by autistics, yet even healthy young children (*ca.* 6 years old) learned this. The authors suggest that part of the difference in behaviour is occurring because autistics have not developed social 'fairness' rules that most healthy individuals have internalized through repeated social interactions. The veracity of this claim would be clarified with measurements of neural activity.

A second example of the need for convergence comes from Knutson's laboratory (Bjork *et al.* 2004) who replicated the paradigm of Knutson *et al.* (2003) using 12 adolescents (ages 12–17 years) and 12 young adults (ages 22–28 years) as subjects. Rewards for the correctly cued colour choice were $0.20, $1 or $5, and choices were designed so that subjects were correct 70% of the time. Gain acquisition in both age groups similarly activated the MPFC. Interestingly, while anticipation of gains activated the ventral striatum in both groups, adolescents had a significantly

lower average BOLD signal than young adults for the same sized reward. These data indicate that one reason adolescents may engage in risky behaviours is to compensate for hypoactive reward activity in their brains. It also suggests that to fully understand anticipation and consumption of rewards, one cannot only study young healthy adults.

The above review of the neuroeconomics literature is, by necessity, truncated and subject to my own biases. Other discussions of the neuroeconomics literature and methodology can be found in Camerer (2003), Camerer *et al.* (2004), the 2002 special issue of *Neuron* (Cohen & Blum 2002), a special issue of *Games and economic behavior* (in the press) and the book by Glimcher (2003).

5. NEUROECONOMICS AND THE LAW

One of most important areas that neuroeconomics can contribute to is the law. Laws (or more generally institutions as defined by Douglass North (1990)) specify the 'rules of the game'; yet not everyone follows these rules. Neuroeconomics experiments that vary the 'laws' and allow subjects to make choices under several legal regimes could be an important step towards better public policy. Such experiments could provide a deep understanding into the usefulness of laws that are either 'carrots' or 'sticks'. For example, when an action results in a harsh punishment (e.g. experimentally, a decrement of money), why do some subjects still choose to do this? What drives such behaviour? How much of it can be traced to nature and nurture? Do known criminals have different neural activity than non-criminals? The number of interesting questions is manifold. The late Margaret Gruter, of the Gruter Institute for Law and Behavioral Research, called this field 'neurojurisprudence'. The economic part is important experimentally because it allows the imposition of acceptable and valued rewards and punishments for behaviours in an experimental setting.

A specific legal example is property crime. Property crime (larceny) is little impacted by most punishments, an increased likelihood of detection, or the provision of presumed alternative leisure activities such as nighttime basketball (Zak 2000). More effective laws might be designed if the neural activation associated with obtaining property illegally, but risking punishment, were characterized. This is straightforward to do in a neuroeconomic experiment using, for example, the 'power-to-take' game (Bosman & van Winden 2002). Neural activity can be measured as rewards and punishments are varied to determine why most punishments fail to deter larceny, and to search for those that are likely to work. The neural activity of larcenists could be compared in this experiment with non-criminals to understand recidivism. In addition, humans appear to have a strong sense of ownership of physical property. Behaviourally, people value an item more when they possess it than when they do not (Camerer 2003). This suggests that people might pay more to protect property than the expected loss associated with its expropriation. There may be neural clues to this behaviour that might suggest why individuals may not want to trade off a given amount of theft for less police protection and lower taxes. This is one example of neuroeconomic–neurojurisprudence complementarity, but many more surely exist. Note that there are a host of important technical and ethical issues that this example opens up, including using averaged brain data to determine policy, using brain-scanning data to identify criminals, appropriate statistical thresholds to determine if something has been demonstrated, etc. The reader is referred to the discussion of these topics by Goodenough & Prehn (2004) and Greene & Cohen (2004) in this issue.

Another transdisciplinary field that also impacts questions of law is neuroethics (Greene & Haidt 2002; Moreno 2003). The notion that some behaviours are almost universally considered wrong is among the first issues that neuroethicists have studied. Greene *et al.* (2001) showed, using fMRI, that personal moral dilemmas (e.g. whether it is morally acceptable to personally kill one person to save five others from certain death) activated cortical areas associated with social cognition, including the medial OFC (BA9/10), posterior cingulate (BA39) and angular gyrus (BA39). Interestingly, regions associated with working memory (BA46, BA7/40) exhibited reduced neural activity during personal moral dilemmas. A legal implication of this research is that laws designed to prohibit personal moral violations must activate brain regions associated with understanding others to be effective.

In fMRI research similar to that of Greene *et al.* (2001) (though with substantially different control tasks), Moll *et al.* (2002) found that moral judgements are associated with significant BOLD signals in the medial OFC, as well as in the temporal pole (BA38) and superior temporal sulcus (BA21/22). This provides support for the role of emotions in moral judgments. Both the Greene and Moll studies could be extended using neuroeconomic methods (e.g. using monetary rewards and punishments) so that subjects' choices have weight and their attention is consistently focused on the task. Further, by varying the 'costs' of immoral behaviour, the robustness of moral disgust could be probed.

6. CONCLUSION

The nineteenth century economist Thorstein Veblen wrote in 1898 that 'Economics, properly understood, is simply a branch of biology'. Human beings are a biological species doing what every other species seeks to do: survive and reproduce (albeit with a larger brain than most other species). These activities require that choices be made to acquire resources, i.e. to process environmental signals, value alternatives and chose among them. Resource acquisition may also require that we interact with other humans, sometimes strategically. Neuroeconomics provides a unified framework to measure neurophysiological activity during the process of choice, and in doing so opens a window into human nature.

REFERENCES

Adolphs, R. 2003 Cognitive neuroscience of human social behaviour *Nature Rev. Neurosci.* **4**, 165–178.

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E. & Breiter, H. C. 2001 Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron* **32**, 537–551.

Barraclough, D. J., Conroy, M. L. & Lee, D. 2004 Prefrontal cortex and decision making in a mixed-strategy game. *Nature Neurosci* **7**, 404–410.

Berg, J., Dickhaut, J. & McCabe, K. 1995 Trust, reciprocity, and social history. *Games Econ. Behav.* **10**, 122–142.

Bjork, J. M., Knutson, B., Fong, G. W., Caggiano, D. M., Bennett, S. M. & Hommer, D. W. 2004 Incentive-elicited brain activation in adolescents: similarities and differences from young adults. *J. Neurosci* **24**, 1793–1802.

Bosman, R. & van Winden, F. 2002 Emotional hazard in a power-to-take game experiment. *Econ. J.* **112**, 147–169.

Bowles, S. & Gintis, H. 2003 Prosocial emotions. Santa Fe Institute working paper no. 02-07-028.

Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. 2003 The evolution of altruistic punishment. *Proc. Natl Acad. Sci. USA* **100**, 3531–3535.

Breiter, H. C., Aharon, I., Kahneman, D., Anders, D. & Shizgal, P. 2001 Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* **30**, 619–639.

Brosnan, S. F. & de Waal, F. B. M. 2003 Monkeys reject unequal pay. *Nature* **425**, 297–299.

Buckner, R. L. 2003 The hemodynamic inverse problem: making inferences about neural activity from MRI signals. *Proc. Natl Acad. Sci. USA* **100**, 2177–2179.

Camerer, C. F. 2003 Strategizing in the brain. *Science* **300**, 1673–1675.

Camerer, C. F., Loewenstein, G. & Prelec, D. 2004 Neuroeconomics: how neuroscience can inform economics. *J. Economic Lit.* (In the press.)

Cohen, J. D. & Blum, K. I. 2002 Reward and decision. *Neuron* **36**, 193–198.

Damasio, A. R. 1994 *Descartes' error: emotion, reason, and the human brain*. New York: Avon Books.

Dayan, P. & Balleine, B. W. 2002 Reward, motivation, and reinforcement learning. *Neuron* **36**, 285–298.

Dickhaut, J., McCabe, K., Nagode, J. C., Rustichini, A., Smith, K. & Pardo, J. V. 2003 The impact of the certainty context on the process of choice. *Proc. Natl Acad. Sci. USA* **100**, 3536–3541.

Fehr, E. & Rockenbach, B. 2003 Detrimental effects of sanctions on human altruism. *Nature* **422**, 137–140.

Frith, U. & Frith, C. D. 2003 Development and neurophysiology of mentalizing. *Phil. Trans. R. Soc. Lond. B* **358**, 459–473. (doi:10.1098/rstb.2002.1218)

Gheslin, M. & Landa, J. T. 1999 The emerging discipline of bioeconomics: aims and scope of the journal of bioeconomics *J. Bioecon.* **1**, 5–12.

Garris, P. A., Kilpatrick, M., Bunin, M. A., Michael, D., Walker, O. D. & Wightman, R. M. 1999 Dissociation of dopamine release in the nucleus accumbens from intracranial self-stimulation. *Nature* **398**, 67–69.

Glimcher, P. W. 2003 *Decisions, uncertainty, and the brain: the science of neuroeconomics*. Cambridge, MA: MIT Press.

Glimcher, P. W., Dorris, M. C., Bayer, H. M. & Lau, B. 2004 Physiologic utility theory and the neuroeconomics of choice. *Games Econ. Behav*. (In the press.)

Goodenough, O. R. & Prehn, K. 2004 A neuroscientific approach to normative judgment in law and justice. *Phil. Trans. R. Soc. Lond.* B **359**, 1709–1726. (doi:10.1098/rstb.2004.1552)

Greene, J. & Haidt, J. 2002 How (and where) does moral judgment work? *Trends Cogn. Sci.* **6**, 517–523.

Greene, J. & Cohen, J. 2004 For the law, neuroscience changes nothing and everything. *Phil. Trans. R. Soc. Lond.* B **359**, 1775–1785. (doi:10.1098/rstb.2004.1546)

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. 2001 An fMRI investigation of emotional engagement in moral judgment. *Science* **293**, 2105–2108. (doi: 10.1126/science.1062872)

Hill, E. & Sally, D. 2003 Dilemmas and bargains: autism, theory-of-mind, cooperation and fairness. Working paper, University College London.

Hirshleifer, J. 1985 The expanding domain of economics. *Am. Econ. Rev.* **75**, 53–68.

Hirshleifer, J. & Zak, P. J. 2004 The bioeconomics of social behavior: introduction. *J. Bioecon.* **6**, 1–2.

Kahn, I., Yeshurun, Y., Rotshtein, P., Fried, I., Ben-Bashat, D. & Hendler, T. 2002 The role of the amygdala in signaling prospective outcome of choice. *Neuron* **33**, 983–994.

Kahneman, D. 2004 Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* **93**, 1449–1475.

Kangarlu, A., Burgess, R. E., Zhu, H., Nakayama, T., Hamlin, R. L., Abduljalil, A. M. & Robitaille, P. M. 1999 Cognitive, cardiac, and physiological safety studies in ultra high field magnetic resonance imaging. *Magn. Reson. Imag.* **17**, 1407–1416.

Knutson, B. & Peterson, D. 2004 Neurally reconstructing expected utility. *Games Econ. Behav.* (In the press.)

Knutson, B., Fong, G. W., Adams, C. M., Varner, J. & Hommer, D. 2001 Dissociation of reward anticipation and outcome with event-related fMRI. *NeuroReport* **12**, 3683–3687.

Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. S. & Hommer, D. 2003 A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *NeuroImage* **18**, 263–272.

Laibson, D., Repetto, A. & Tobacman, J. 1998 Self-control and savings for retirement. *Brook. Papers Econ. Act.* **1**, 91–196.

Lo, A. W. & Repin, D. 2002 The psychophysiology of real-time financial risk processing. *J. Cogn. Neurosci.* **14**, 323–339.

McCabe, K., Houser, D., Ryan, L., Smith, V. & Trouard, T. 2001 A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl Acad. Sci. USA* **98**, 11 832–11 835.

McClure, S. M., Laibson, D. I., Loewenstein, G. & Cohen, J. D. 2004 Separate neural systems value immediate and delayed monetary rewards. *Science* **306**, 2105–2108. (doi: 10.1126/science.1100907)

Mobbs, D., Greicius, M. D., Abdel-Azim, E., Menon, V. & Reiss, A. L. 2003 Humor modulates the mesolimbic reward centers. *Neuron* **40**, 1041–1048.

Moll, J., de Oliveira-Souza, R., Bramati, I. E. & Grafman, J. 2002 Functional networks in emotional moral and nonmoral judgments. *NeuroImage* **16**, 696–703.

Montague, R. P. & Berns, G. S. 2002 Neural economics and the biological substrates of valuation. *Neuron* **36**, 265–284.

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., Wiest, M. C., Karpov, I., King, R. D., Apple, N. & Fisher, R. E. 2002 Hyperscanning: simultaneous fMRI during linked social interactions. *NeuroImage* **16**, 1159–1164.

Moreno, J. D. 2003 Neuroethics: an agenda for neuroscience and society. *Nature Rev. Neurosci.* **4**, 149–153.

Nelson, A. J., Heeger, D. J., McCabe, K., Houser, D., Zak, P. & Glimcher, P. W. 2004 Expected utility provides a model for choice behavior and brain activation in humans. Abstract No. 20.12. Society for Neuroscience.

North, D. 1990 *Institutions, institutional change and economic performance*. Cambridge University Press.

Platt, M. L. & Glimcher, P. W. 1999 Neural correlates of decision variables in parietal cortex. *Nature* **400**, 233–238.

Riedel, G., Platt, B. & Micheau, J. 2003 Glutamate receptor function in learning and memory. *Behav. Brain Res.* **140**, 1–47.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S. & Kilts, C. D. 2002 A neural basis for social cooperation. *Neuron* **35**, 395–405.

Robson, A. J. 2001 Why would nature give individuals utility functions? *J. Polit. Econ.* **109**, 900–914.
Rustichini, A., Dickhaut, J., Ghirardato, P, Smith, P. & Glimcher, P.W. 2004 Expected utility provides a model for choice behavior and brain activation in humans. Abstract No. 20. 12. Society for Neuroscience.
Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. 2003 The neural basis of economic decision-making in the ultimatum game. *Science* **300**, 1755–1758. (doi: 10.1126/science.1082976)
Schultz, W., Dayan, P. & Montague, P. R. 1997 A neural substrate of prediction and reward. *Science* **275**, 1593–1599. (doi: 10.1126/science.275.5306.1593)
Smith, K., Dickhaut, J., McCabe, K. & Pardo, J. 2002 Neuronal substrates for choice under ambiguity, risk, certainty, gains, and losses. *Mngmt Sci.* **48**, 711–718.
Smith, V. 1998 The two faces of Adam Smith. *South. Econ. J.* **65**, 1–29.
Sugrue, L. P., Corrado, G. S. & Newsome, W. T. 2004 Matching behavior and the representation of value in the parietal cortex. *Science* **304**, 1782–1787. (doi: 10.1126/science.1094765)
Zak, P. J. 2000 Larceny. *Econ. Govern.* **1**, 157–179.
Zak, P. J. 2002 Genetics, family structure, and economic growth *J. Evol. Econ.* **12**, 343–365.
Zak, P. J. & Knack, S. 2001 Trust and growth. *Econ. J.* **111**, 295–321.
Zak, P. J. & Denzau, A. 2001 Economics is an evolutionary science. In *Evolutionary approaches in the behavioral sciences: toward a better understanding of human nature* (ed. A. Somit & S. Peterson), pp. 31–65. New York: JAI Press.
Zak, P. J. & Park, K.-W. 2002 Population genetics and economic growth. *J. Bioecon.* **4**, 1–37.
Zak, P. J., Kurzban, R. & Matzner, W. 2004 The neurobiology of trust. *Ann. NY Acad. Sci.* **1032**. (In the press.)

GLOSSARY

ACC: anterior cingulate cortex
BOLD: blood oxygen-level dependent
DLPFC: dorsolateral prefrontal cortex
DM1: decision maker 1
DM2: decision maker 2
EEG: electroencephalogram
ERP: evoked response potential
fMRI: functional magnetic resonance imaging
LIP: lateral intraparietal
LTP: long-term potentiation
MPFC: mesial prefrontal cortex
OFC: orbitofrontal cortex
OFS: orbitofrontal–striatal
OT: oxytocin
PD: Prisoner's Dilemma
PET: positron emission tomography
SLEA: sublenticular extended amygdala
TMS: transcranial magnetic stimulation