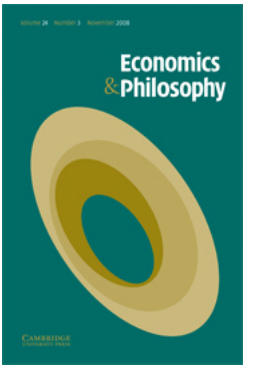

NEUROECONOMICS: A CRITICAL RECONSIDERATION

Glenn W. Harrison

Economics and Philosophy / Volume 24 / Special Issue 03 / November 2008, pp 303 - 344
DOI: 10.1017/S0266267108002009, Published online: 05 November 2008

Link to this article: http://journals.cambridge.org/abstract_S0266267108002009

How to cite this article:
Glenn W. Harrison (2008). NEUROECONOMICS: A CRITICAL RECONSIDERATION. Economics and Philosophy, 24, pp 303-344 doi:10.1017/S0266267108002009

Economics and Philosophy, **24** (2008) 303–344
doi:10.1017/S0266267108002009

NEUROECONOMICS: A CRITICAL RECONSIDERATION

Glenn W. Harrison*
University of Central Florida

Understanding more about how the brain functions *should* help us understand economic behaviour. But some would have us believe that it has done this already, and that insights from neuroscience have already provided insights in economics that we would not otherwise have. Much of this is just academic marketing hype, and to get down to substantive issues we need to identify that fluff for what it is. After we clear away the distractions, what is left? The answer is that a lot is left, but it is still all potential. That is not a bad thing, or a reason to stop the effort, but it does point to the need for a serious reconsideration of what neuroeconomics is and what passes for explanation in this literature. I argue that neuroeconomics can be a valuable field, but not the way it is being developed and "sold" now. The same is true more generally of behavioural economics, which shares many of the methodological flaws of neuroeconomics.

Understanding more about how the brain functions *should* help us understand economic behaviour. But some would have us believe that it has done this already, and that insights from neuroscience have already provided insights in economics that we would not otherwise have. Much of this is just academic marketing hype, and to get down to substantive issues we need to identify that fluff for what it is. After we clear that

* Department of Economics, College of Business Administration, University of Central Florida, USA, and Durham Business School, Durham University, UK (part-time). Email contact: gharrison@research.bus.ucf.edu. Thanks to the U.S. National Science Foundation for research support under grants NSF/HSD 0527675 and NSF/SES 0616746, to Michael Caputo, John Dickhaut, Mark Dickie, Milton Heifetz, Richard Hofler, Julian Jamison and Elisabet Rutström for helpful discussions, and to John Hey and Giacomo Bonanno for the invitation to be a piñata.

away as a distraction, what is left? The answer is that a lot is left, but it is still all potential. That is not a bad thing, or a reason to stop the effort, but it does point to the need for a serious reconsideration of what neuroeconomics is and what passes for explanation in this literature. I argue that neuroeconomics can be a valuable field, but not the way it is being developed and "sold" now. The same is true more generally of behavioural economics, which shares many of the methodological flaws of neuroeconomics.[1]

Epistemology is the study of knowledge claims in philosophy. My complaints are with the epistemological basis of neuroeconomics research as it has been practised. Most of the claims appear to build on poor experimental and statistical foundations, to the point where we end up with a mixture of ad hoc statistical evidence and story-telling, presented as truth or knowledge. In many cases pre-confused experimental designs have simply had neural data tacked on for academic novelty value. Again, this is fine as long as we do not present it as more than it is. But once we deal with what we have before us, we would like to ask the deeper question about whether neuroeconomics can ever add knowledge to economics. Some have already claimed that it cannot, and those claims have to be evaluated as well.

In section 1 I discuss some marketing bloopers of the neuroeconomics literature, since this promotional material is just a distraction from the main issues but is widely cited.[2] If this is taken for real science, then it is already "game over" in terms of the deeper contribution that neuroeconomics might, and should, make (Rubinstein 2006). In section 2 I discuss some general concerns with applications of neuroeconomics, returning more precisely to some themes emerging in shadowy form in section 1. In section 3 I consider some of the debates over the validity of data from

[1] Excellent, uncritical overviews of the methods and jargon of neuroeconomics are provided by Camerer *et al*. (2004, 2005). The methods involve more than just brain imaging, as they explain, although that method is the poster boy of the field. Some, such as Ross (2005: 322) define neuroeconomics partly in terms of a research program, perhaps best associated with Glimcher (2003) and Montague and Berns (2002), by way of Marr (1982), that uses economics principles to explain the evolution of neural behavior in animals. This program is reconsidered in § 4. Ross (2005) offers a sustained enquiry into the relationship between economics and the philosophy of cognitive science.

[2] I generally ignore the editorial comments by non-economists that accompany reports in journals such as *Science* or *Nature*, although they often have lines that should cause any economist to wince. Camerer (2008) now argues that some of these "early neuroeconomic papers should be read as if they are speculative grant proposals which conjecture what might be learned from studies which take advantage of technological advances" rather than as "logical conclusions derived from mathematical analyses". I do not see how anyone would have ever read them as the latter, since they are always grounded in significant empiricism and inferential regularities rather than mathematical proof. Nor do I understand how one can claim a mulligan in scientific discourse.

neuroeconomics collected in Caplin and Schotter (2008). Section 4 offers a broader framework to think about the contributions that psychology and neuroscience can make to understanding economic behaviour, responding constructively to the concern that Smith (2007: ch.14) expressed about the role of neuroeconomics. By broadening the questions we ask, rather than just trying to answer the same questions with new toys, I believe we can see where neuroeconomics will eventually make a significant contribution to our scientific knowledge about economic behaviour. For now, that promise remains just that.

1. MARKETING BLOOPERS IN THE SELLING OF NEUROECONOMICS

Camerer, Loewenstein and Prelec (2005) (CLP) provide a dramatic statement of "how neuroscience can inform economics". They undertake a valuable counterfactual thought experiment: how would economics have evolved if it had been informed at the outset by what we now know from neuroscience? This is not just Monday Morning Quarterbacking, but exactly the kind of meta-analysis of the history of thought in economics that philosophy teaches us to be a valuable methodological tool. So, primed to the purpose of their task, what a disappointment is in store! They claim that neuroscience "... points to an entirely new set of constructs to understand economic decision-making" (p. 10). In virtually all cases, however, the claims are just recycled from earlier work in judgement and decision-making, and do not represent insights gained from neuroeconomics as such. But are they, indeed, insights? Let's see.

1.1 Fundamental new concepts?

CLP (p. 32) claim that understanding about how the brain works

> ... challenges some of our fundamental assumptions about how people differ from one-another when it comes to economic behavior. Economists currently classify individuals on such dimensions as "time preference," "risk preference," and "altruism." These are seen as characteristics that are *stable within an individual over time and consistent across activities*; someone who is risk-seeking in one domain is expected to be risk-seeking in other domains as well. But empirical evidence shows that risk-taking, time discounting, and altruism are very weakly correlated or uncorrelated across situations. This inconsistency results in part from the fact that preferences are state-contingent (and that people may not recognize the state-contingency, which – if they did – would trigger overrides that impose more consistency than observed). But it also may point to fundamental problems with the constructs that we use to define how people differ from each other. [italics added]

So we learn from neuroeconomics that we should, in general, allow state-dependent preferences? This approach would allow a person to have

different discount rates for saving, flossing of teeth, dieting, and the decision to get a tattoo, to take their examples.

Of course, economists have known this for decades, and book-length treatments by Hirshleifer and Riley (1992) and Chambers and Quiggin (2000) review the massive literature. Andersen *et al*. (2008b) illustrate how one can apply it to determine if preferences are stable over time, and the subtlety of putting operational structure on that question even in a controlled experimental setting.

Whether or not we use a state-dependent approach in analysis is a separate matter. The extreme alternative, and no straw man, is presented by Stigler and Becker (1977). They are clearly proposing the view that assuming that preferences are stable, and common across individuals, is simply a more useful approach. Judgement day on that hypothesis can wait for our purposes, although there is a behaviourist undercurrent in much of the neuroeconomics literature that presumes that judgement has been made.

But consider the evidence provided to support the claim that we have a challenge to fundamental concepts. "For example, whether a person smokes is sometimes taken as a crude proxy for low rates of time discounting, in studies of educational investment or savings" (p. 32). The use of a crude proxy, no doubt openly acknowledged as same, in an empirical effort, is somehow evidence that we need a new *concept* of time preferences? I miss the point. The remaining examples are no less anecdotal, and simply irrelevant to the argument (e.g. flossing being correlated with feeding parking meters).

Camerer, Loewenstein and Prelec (2004: 563) make the same point with a dripping of sarcasm that plays well in plenary lectures, but can be waved aside by any student of economics:

> Another example suggests how concepts of economics can be even wider off the mark by neglecting the nature of biological state-dependence: Nobody chooses to fall asleep at the wheel while driving. Of course, an imaginative rational-choice economist – or a satirist – could posit a tradeoff between 'sleep utility' and 'risk of plowing into a tree utility' and infer that a dead sleeper must have had higher u(sleep) than u(plowing into a tree). But this 'explanation' is just tautology.

And lousy, sucker-punch, economics.

1.2 Utility for money?

CLP (p. 35) argue that

> . . . neuroscience can point out commonalities between categories that had been viewed as distinct. An example of this with important implications for

> economics is the utility for money. The canonical economic model assumes that the utility for money is indirect – i.e., that money is a mere counter, only valued for the goods and services it can procure. Thus, standard economics would view, say, the pleasure from food or cocaine and the "pleasure" from obtaining money as two totally different phenomena. Neural evidence suggests, however, [. . .] that money provides direct reinforcement.

Huh?

First, there are several popular models in economics in which money is an argument of the utility function. Some economists grump about those models, but it is more a matter of modelling taste than a clear, defining line between "the one, true model of economics" and "reduced forms used in the heat of the expositional moment". Important parts of economics written in terms of utility functions defined directly over payoffs include behavioural game theory (Camerer 2003), auction theory, contract theory, and so on. The *opening* sentence in Pratt's (1964) classic defining measures of risk aversion is pretty clear on the matter: "Let u(x) be a utility function for money." If that is not canonical enough for you, then you need to re-load your canon.

Second, it does not take much head scratching to envisage a two-stage, nested utility function in which money enters at some top level, and consumption of non-money goods are purchased with that money. It is simply self-serving formalism, common from the behavioural literature, to construct a straw-man model to attack. Thus, in a modern text on contract theory, Bolton and Dewatripont (2005: 4) get on with their work without formal semantics on the issue:

> Let us denote the employer's utility function as $U(l, t)$ where l is the quantity of employee time the employer has acquired and t denotes the quantity of "money" – or equivalently the "output" that this money can buy [footnote omitted] – that he has at his disposal. Similarly, employee utility is $u(l, t)$, where l is the quantity of time the employee has kept for herself and t is the quantity of money that she has at her disposal.

The omitted footnote neatly clarifies the obvious to any practising economist: "Indeed, the utility of money here reflects the utility derived from the consumption of a composite good that can be purchased with money." So if the same reward circuits of the brain fire when subjects get money or beer, that is fine with any number of representations of the utility function in economics.[3]

[3] The same point holds true for money allocated to one's self or allocated to some public good, of course, so there should be no surprise when Harbaugh *et al*. (2007) show that the same reward centres light up for both allocations.

1.3 Wants are not likes?

CLP (p.37) next take aim at an alleged assumption in economics:

> Economists usually view behaviour as a search for pleasure (or, equivalently, escape from pain). The subfield of welfare economics, and the entire ability of economists to make normative statements, is premised on the idea that giving people what they want makes them better off. But, there is considerable evidence from neuroscience and other areas of psychology that the motivation to take an action is not always closely tied to hedonic consequences. Berridge (1996) argues that decision making involves the interaction of two separate, though overlapping systems, one responsible for pleasure and pain (the "liking" system), and the other for motivation (the "wanting" system). This challenges the fundamental supposition in economics that one only strives to obtain what one likes.

No, this is not what economics assumes at all. We say that choices reveal preferences, on a good inferential day, which is not even close. Binmore (2007a: 111) explains the methodological difference between these points of view well:

> To speak of utility is to raise the ghost of a dead theory. Victorian economists thought of utility as measuring how much pleasure or pain a person feels. Nobody doubts that our feelings influence the decisions we make, but the time has long gone when anybody thought that a simple model of a mental utility generator is capable of capturing the complex mental process that swings into action when a human being makes a choice. The modern theory of utility has therefore abandoned the idea that a util can be interpreted as one unit more or less of pleasure or pain. One of these days, psychologists [neuroeconomists?] will doubtless come up with a workable theory of what goes on in our brains when we decide something. In the interim, economists get by with *no theory at all* of why people choose one thing rather than another. The modern theory of utility makes no attempt to explain choice behavior. It assumes that we already know what people choose in some situations and uses this data to deduce what they will chose in others – on the assumption that their behavior is consistent.

To non-economist readers, the "modern" theory in question began with Samuelson (1938).

1.4 Domain-specific expertise?

CLP (p. 33) argue that "Economics implicitly assumes that people have general cognitive capabilities that can be applied to any type of problem and, hence, that people will perform equivalently on problems that have similar structure." Really? I thought we had a pretty good working theory of human capital and compensating wage differentials for between-subject comparisons.

Is this also true, however, interpreted as applying on a within-subjects basis? If somebody is presented with an abstract logic puzzle, such as the renowned Wason Selection Task they reference (Wason and Johnson-Laird 1972), and represents it differently than when it is presented as a concrete puzzle, it is not just idle semantics to say that they are solving different problems. One can argue that it is a different task when the subject perceives it differently, and then the notion of "similarity of structure" becomes a rich research question in search of a metric.[4]

Of course, one does not want to take that argument too far, or else it guts theory of any generality. The real point of this example is that people tend to perform better at this task when it is presented with concrete, naturally occurring referents, as distinct from an abstract representation.[5] Moreover, there is important evidence that exposure to realistic instances fails to transfer to abstract instances (Johnson-Laird, Legrenzi and Legrenzi 1972). These remain important findings, and tell us that the ability to correctly apply principles of logical reasoning depends on syntax *and* semantics.

The contextual nature of performance is important, but it is a mistake to again think that it is intrinsic to economic reasoning, as distinct from something that is sometimes, or often, assumed away for convenience.

The final example is the Tower of Hanoi puzzle, extensively studied by cognitive psychologists (e.g. Hayes and Simon 1974) and more recently by economists (McDaniel and Rutström 2001) in some fascinating experiments.[6]

[4] One does not need to look far to find such metrics, and their relevance for economics: Tversky (1969, 1977), Rubinstein (1988) and Leland (1994).

[5] This qualitative result was first identified by Wason and Shapiro (1971). The psychology literature quickly noted that performance on the abstract and concrete versions varies significantly from sample to sample, and from variations in task familiarity (e.g. Gilhooly and Falconer 1974: 358ff.). This does not invalidate the core point, but adds a caution against unqualified statements.

[6] The Tower of Hanoi involves $n \geq 3$ disks piled onto one of $k \geq 3$ pegs, with higher disks being smaller. Call this the initial state, and assume $k = 3$, and that $n > k-1$ to make the task interesting. The goal is to move all of the disks to peg 3. The constraints are that only one disk may be moved at a time, and no disk may ever lie under a bigger disk. The objective is to reach the goal state in the least number of moves. The "trick" to solving the Tower of Hanoi is to use backwards induction: visualize the final, goal state and use the constraints to figure out what the penultimate state must have looked like (viz., the tiny disk on the top of peg 3 in the goal state would have to be on peg 1 or peg 2 by itself). Then work back from that penultimate state, again respecting the constraints (viz., the second smallest disk on peg 3 in the goal state would have to be on whichever of peg 1 or peg 2 the smallest disk is *not* on). One more step in reverse and the essential logic should be clear (viz., in order for the third largest disk on peg 3 to be off peg 3, one of peg 1 or peg 2 will have to be cleared, so the smallest disk should be on top of the second smallest disk).

Casual observation of students in Montessori classrooms makes it clear how they (eventually) solve the puzzle, when confronted with the initial state. They shockingly violate the constraints and move all the disks to the goal state *en masse*, and then physically work backwards along the lines of the above thought experiment in backwards induction. The critical point here is that they temporarily violate the constraints of the problem in order to solve it "properly".

Contrast this behaviour with the laboratory subjects in McDaniel and Rutström (2001). They were given a computerized version of the game, and told to try to solve it. However, the computerized version did not allow them to violate the constraints. Hence the laboratory subjects were unable to use the classroom Montessori method, by which the student learns the idea of backwards induction by exploring it with physical referents. This is not a design flaw of these lab experiments, but simply one factor to keep in mind when evaluating the behaviour of their subjects. Without the physical analogue of the final goal state being allowed in the experiment, the subject was forced to visualize that state conceptually, and to likewise imagine the penultimate states. Although that might encourage more fundamental conceptual understanding of the idea of backwards induction, if attained, it is quite possible that it posed an insurmountable cognitive burden for some of the experimental subjects.[7]

Harrison and List (2004: 1024) use this example to illustrate one of several defining characteristics of field experiments, in contrast to the typical lab experiment:

> It might be tempting to think of this as just two separate tasks, instead of a real commodity and its abstract analogue. But we believe that this example does identify an important characteristic of commodities in ideal field experiments: the fact that they allow subjects to adopt the representation of the commodity and task that best suits their objective. In other words, the representation of the commodity by the subject is an integral part of how the subject solves the task. One simply cannot untangle them, at least not easily and naturally. This example also illustrates that off-equilibrium states, in which one is not optimizing in terms of the original constrained optimization task, may indeed be critical to the attainment of the equilibrium state. Thus we should be mindful of possible field devices which allow subjects to explore off-equilibrium states, even if those states are ruled out in our null hypotheses.

[7] Dehaene and Changeux (1997) develop an explicit neuronal, computational model of planning behaviour in a closely related task known as the Tower of London. Their model nicely integrates behavioural observations of behaviour in this task, as well as differential behaviour by subjects that have a lesioned prefontal cortex.

So the point is that tasks might look similar, logically or nominally, from the metric of the *equilibrium* prediction of some theory given some assumption about the stock of knowledge the subject has, but not be at all similar when one considers off-equilibrium behaviour or heterogeneity in the stock of knowledge subjects have. It is simply inappropriate to claim that economics has no interest in the latter environments, even if it has had a lot more to say about the former environments. We return to this point later, since it alerts us to the importance of thinking about the process leading to choice, rather than the choice itself. This will be the key to seeing what role neuroeconomics can play.

2. BUT IS IT GOOD ECONOMICS?

So much for the marketing: what are we to make of some of the major claims in the neuroeconomics literature? Is it good, interesting economics? We first consider some general issues to do with samples, statistics and procedures. These are troubling, and need to be made explicit because they are blurred in the existing literature, but in the end not the stuff that we should use to pass judgment on the potential role of neuroeconomics. We then examine three illustrative areas of substantive research that should be of immediate interest to economists: discounting over time, social preferences and trust, and the strategic sense that we expect to see in subjects playing games. The theme to emerge is that many confounds that are known in the experimental economics literature are glossed, and the contribution is to simply tack on some neurological data to help tell the preferred story. This brief review provides a basis for asking, in section 3, if this is just a reflection of a nascent field and we should just be more patient, or if it is fundamental.

2.1 Questions about procedures

Sample sizes for many neuroeconomics studies relying on imaging are small if we count a brain as the unit of analysis.[8] It is common to see studies where the sample size is less than 10, and rarely does one see much more than a dozen or so.[9] To take a recent example, and something of an

[8] Neuroeconomics involves more than imaging, of course. In some cases the samples are quite large, as in Kosfeld *et al*. (2005), where 194 subjects were used.

[9] For example, and spanning several years of funding growth, Delgado *et al*. (2000), Elliott *et al*. (2000), Knutson *et al*. (2000), Dickhaut *et al*. (2003), McClure *et al*. (2004, 2007), Rilling *et al*. (2004), Hsu *et al*. (2005), Harbaugh *et al*. (2007) and Chandrasekhar *et al*. (2008) used 9, 9, 12, 9, 20 (experiment 1) and 14 (experiment 2), 14, 19, 16, 19 and 30 subjects, respectively. One outlier is Lohrenz *et al*. [(2007), who report 54 subjects. It is fair to point out that sample sizes in the early literature in experimental economics were also sometimes small: for example, Friedman, Harrison and Salmon (1984) reported 1 data point per treatment in one of the first generation of experimental asset markets.

outlier at that, Bhatt and Camerer (2005) used 16 subjects, and noted "To experimental social scientists, 16 seems like a small sample. But for most fMRI studies this is usually an adequate sample to establish a result because adding more subjects does not alter the conclusions much" (p. 432; fn.12). One has to wonder how one can possibly draw the final inference without actually doing it for this experiment and set of statistical inferences. There is presumably some good budgetary explanation for only using small numbers of subjects, particularly in an era in which many neuroeconomists now have their own imaging machinery or ready access to same.

But the unit of analysis is not actually the brain: it is a spatial location in the brain firing per unit of time. So we end up with a data set in which a few brains contribute many observations at each point in time, and in a time-series. Now, econometricians know a few things about handling data like this, but what role do those methods play in the analysis? No problem if they are called something else, but the statistical analysis of these neural data is currently a black box in expositions of neuroeconomics research. Bullmore *et al*. (1995) and Rabe-Hesketh *et al*. (1997) provide an excellent exposition of the inner workings of this black box, which is a classic "sausage making factory".

The statistical issues break down into inferences about a single brain, and then a host of new issues coming from inferences over a pooled sample of brains. To understand the significance of the former, here is a list of the estimation methods Rabe-Hesketh *et al*. (1997) walk through in a typical statistical analysis: filtered Fourier transformations of the raw images (p. 217), interpolation to ensure re-alignment to the first image for that subject (p. 218), simple polynomial regression to correct for intra-subject movement (p. 219 and Figure 2), possible corrections for magnetic field inhomgeneity (p. 219), linear regression of signal intensity value (which is an estimate obtained from prior steps) on a dummy to reflect treatments (p. 221), estimation of a response function to reflect haemodynamically mediated delay between stimulus and response (p. 222–5), and *then* we get to the point where we start really opening up the econometrics toolkit to worry about serial correlation and heteroscedasticity on a within-subject basis (p. 226ff.). This list varies from application to application, but is sufficiently representative.

Now open up the inferential can of worms involved in pooling across brains. The clinical need to do this arose from a desire to increase statistical power, and now has become the *conditio sine qua non* of neural comparisons of patients developing symptoms of Alzheimers (so the data become longitudinal). Andreasen *et al*. (1994) famously identified the role of abnormalities of the thalamus in patients with schizophrenic disorders by conducting a meta-analysis of scans of 47 normal brains and 39 brains from known schizophrenics. Techniques for normalizing the brain scans,

discussed below, were applied, and

> an "average brain" was generated for the schizophrenic group and for the normal group. This average brain can be visualized and resampled three-dimensionally in the same manner as an individual MR data set. It has the advantage, however, of providing a concise numeric and visual summary of the group as a whole. (p. 295)

Note that the "summary" in question is not meant to be descriptive: the whole point is that it is to be used inferentially. But there are different ways to normalize brains, as one can imagine (and hope).[10] And it matters for inference. Tisserand *et al*. (2002) examine the impact on inferences about the effect of ageing on regional frontal cortical volumes of using a labor-intensive, manual tracing method, a semi-automatic, partial-brain, volumetric method, and the now-popular whole-brain, voxel-based morphometric methods. They find significant differences across methods, and conclude (p. 667) that

> . . . despite the clear advantages of automatic and voxel-based approaches (quick, perfectly reproducible, applicable to large datasets), the current findings suggest that, at present, the most accurate method is still an anatomically based manual tracing one.

The same point is developed in an important evaluation of "Zen and the art of medical image registration" by Crum *et al*. (2003: 1435):

> VBM (Voxel-Based Morphometry) is undoubtedly a powerful framework that successfully removes the need for expert labour-intensive segmentation but currently replaces it with a complicated problem of interpretation and validation which significantly reduces its efficacy. The validity of current published studies relying on NRR (Non-Rigid Registration) in this way is in most cases limited and in some cases suspect due to indiscriminate application of these poorly understood techniques. Until the technology can be made demonstrably more robust one possible solution is to accept that VBM should only be used as a prompting system to highlight regions of the brain worthy of further analysis by more manually intensive techniques. This approach combines the hypothesis-free advantages of VBM with the benefits of expert manual intervention but is by no means an ideal solution.

So it is not just that there are differences in the methods, with some inferential significance, but that the automatic methods have open issues of validation and interpretation. It is apparent from their earliest statements, such as Friston *et al*. (1995), that these methods, apparently universally

[10] Al Roth offers the joke of an fMRI of a car, consisting of "3 Hummers, 2 BMWs and 4 Minis mapped into normalized Talairach coordinates of a Prius" in lecture notes on "Questions for Neuro Social Scientists" (see http://kuznets.fas.harvard.edu/~aroth/QuestionsForNeuroSocialScientists. ppt).

adopted in the neuroeconomics literature, "have deliberately sacrificed a degree of face validity (the ability to independently gauge the correctness of the approach) to ensure construct validity (the validity of the approach by comparison with other constructs)" (Crum *et al*. 2003: 1426). And this is only one of the statistical concerns about where the "left hand side" of the final analysis comes from in neuroeconomics.[11]

A broader problem is that the statistical modelling of neural data is a sequential mixture of limited-information likelihood methods, cobbled together in a seemingly *ad hoc* manner: MacGyver econometrics. Brain scans do not light up like Xmas trees without lots of modelling assumptions being made. Recognizing this fact is not meant as a way of invalidating the clinical or research goals of the exercise, but an attempt to be sure that we understand the extremely limited extent to which modeling and estimation errors are correctly propagated throughout the chain of inferences. The end result is often a statistical test in which "left hand side" *and* "right hand side" variables are themselves estimates, often from a long chain of estimates, and are treated as if they are data that are known for certain. The overall implication is that one should be concerned that there is likely a significant *understatement* of standard errors on estimate of effects, implying a significant *overstatement* of statistical significant differential activation.

The statistical methods used in many cases are those that have received some acceptance in the neuroscience literature, but those norms and practices may have evolved from clinical needs (e.g. the detection of ischemia, to facilitate the early prevention of a stroke) rather than research needs. It is worth stressing that these issues are debated within the specialist field journals. Thus Crum *et al*. (2003) note clearly that

> . . . the most widely used methods are essentially dumb in that, for a particular registration task, they report only a measure of image similarity which does not allow a judgement of "success" or "failure" to be made. Worse, the magnitude and spatial distribution of errors in NRR are unknown and an understanding of exactly how image-similarity measures and arbitrary transformation models combine in the matching of complex sets of imaged features remains out of reach. (p. 1425) [. . .] these automated approaches are only of value if their performance can be evaluated, preferably in a way that

[11] Quite apart from figuring out what is data and what are estimates (and with what standard errors), one wonders how much of the "line integral" of response reflects activation as the subject learns about the task and thinks about it, and how much reflects the state of mind at the point of choice. Years of scholarship in models of learning in games has reminded us of the dangers of confounding substantive models with simplistic statistical modelling when we have panels of time-series, as sharply demonstrated by Wilcox (2006) for the "reinforcement learning model". So it is impossible to know what to make of efforts by Lohrenz *et al*. (2007) to discriminate between experiential and fictitive (counter-factual) reinforcement learning models.

> allows correspondence error to be propagated through subsequent statistical analysis. This is currently not the case. . . (p. 1434)

Constructive statistical approaches that take more of a "full information" approach are also *starting* to be developed in the specialist literature, such as Bowman *et al*. (2008) and Wong *et al*. (2008), but none of these concerns seem to filter through to qualify the conclusions of the neuroeconomics literature.

A serious issue arises, however, from one of the dark secrets of this field: no neural data is apparently ever made public. For some years requests were made to prominent authors in neuroeconomics, particularly economists with significant and well-earned reputations in experimental economics. Not one has provided data. In some cases the economist could not get the neuroscience co-author to release data, perhaps reflecting traditions in that field.[12] In any event, one hopes this unfortunate professional habit ends quickly, particularly for government-funded research from tax dollars. No trips to Stockholm, presumably, until statistical methods can be replicated, extended or evaluated with alternatives.

The procedures of many neuroeconomics tasks seem to gloss some of the norms that have evolved, for good or purely historical reason, in experimental economics. For example, Knoch *et al*. (2006) studied behaviour in an Ultimatum Bargaining game in which subjects' brains were zapped by "low-frequency repetitive transcranial magnetic stimulation" for 15 minutes prior to making their choices. Quite apart from the ethics of inflicting such an electrical tsunami on subjects, raised by Jones (2007) and responded to by Knoch *et al*. (2007), the study used deception to lead subjects to think that human subjects were actually providing offers.[13] The actual offers made were originally made by humans, but in a prior experiment, and the *aggregate* distribution from that prior distribution was used to generate the distribution given to each subject in the main experiment. Thus the distribution actually received would have less variance than the expected predictive distribution of the subjects. The same deception was employed in studies of behaviour in simple games by Sanfey *et al*. (2003) and Rilling *et al*. (2004), and in their case the deception may have confounded inferences about whether the "theory of mind"

[12] At the time, the objective was to evaluate some of the statistical methods used with the aid of a famous neurosurgeon and a specialist in tomographic reconstruction. Now, the objective is just to identify scholars willing to make their data open for examination. The size of the data involved is understood, but is trivial in these days of portable USB drives. In many cases the economist in question gladly provided any "behavioural" non-image data, although there are several prominent publications for which even those data are being withheld "pending further analysis and publications".

[13] This is explained in the online Supplementary Materials provided, and is not mentioned in the main article.

regions of the brain are activated. These are the sorts of cheap tricks in design that experimental economists like to avoid.

A final procedural feature of many neuroeconomics studies is the exceedingly cryptic manner in which things are explained. The neuroscience jargon is not the issue: presumably that sub-field has an acronym-rich semantic structure because they find it efficient. Instead, basic behavioural procedures and statistical tests are minimally explained. Many of the journals in question, particularly the prominent general science journals, have severe page restrictions. But in an era of web appendices, that cannot be decisive. In many cases it is simply impossible to figure out, without an extraordinary amount of careful reading, exactly what was done in the *economics* parts of the experiment and statistical analysis. One wonders how incestuous and unquestioning the refereeing process has become in certain journals.

2.2 Discounting behaviour

The experimental study of time discounting has had a messy history, and procedures have evolved to mitigate concerns with previous studies. The first generation of studies used hypothetical tasks, varying principals, and cognitively difficult ways of eliciting valuations from subjects. The second generation used real rewards and simplified the task in various ways to separate out subjective discount rates from the cognitive burden of evaluating alternatives that differed in several dimensions (principal, front end delay, time horizon). The third generation elicited risk attitudes and discount rates jointly, to allow the latter to be inferred as the rate at which time-dated *utility* streams were being valued rather than as the rate at which time-dated *monetary* streams were being valued; see Andersen *et al*. (2008a) for a review of the literature.

A key procedural issue has been the use of a front end delay on the earlier option. That is, if the subject is asked to choose between an amount of money m in d days and an amount of money M in D days, where $M > m$ and $D > d$, does d refer to the current experimental session ($d = 0$) or some time in the future ($d > 0$)? The reason that this procedural matter assumes such importance is because of the competing explanations for apparently huge discount rates when there is no front end delay ($d = 0$). By "huge" we mean off the charts: this is a literature that earnestly catalogues annual discount rates in the hundreds or thousands when there is no front end delay.

One explanation for this outcome is that behaviour is consistent with "hyerbolicky" discounting, an expression which conjoins at the hip various families of hyperbolic, generalized hyperbolic and quasi-hyperbolic discounting. The most popular exemplar is quasi-hyperbolic discounting, which posits that the individual has one extremely high rate

of time preference for the present versus the future (i.e. maintaining $d = 0$) and a much lower rate of time preference for time-dated choices in the future (i.e. maintaining $d > 0$). This hypothesis suggests to McClure *et al*. (2004, 2007) that there could be different parts of the brain activated when considering $d = 0$ choices and $d > 0$ choices. They do not claim that it is *necessary* for there to be different parts activated for the quasi-hyperbolic specification to be supported, just that it would be consistent with that specification if there was such differential brain activity.

An alternative explanation, recognized variously in comments in the earlier literature, and first stated clearly by Coller and Williams (1999), is that the evidence is an experimental artefact of the credibility and transactions costs of the subject receiving money at the session rather than some other time in the future. This artefact may be what proponents of hyperbolicky specifications mean by a "passion for the present", but it is distinct from any notions of time preference deriving solely from delay in consumption.[14] Again, however, it could be that the parts of the brain that light up when doing discounting calculations are different from the parts of the brain that evaluate immediate rewards. The concept of the counterfactual of "the bigger catch we might get tomorrow if we rest today" presumably evolved in our brains separately from the visceral senses that motivate us to tuck in today to tasty food when it is available.

McClure *et al*. (2004) provide evidence that when subjects face decisions in which "money today" is involved, compared with decisions in which "money in the future" is also involved, different parts of the brain light up. For money today, the same regions identified for "rewards" were differentially activated (ventral striatum, medial orbitofrontal cortex, and median prefontal cortex).[15] However, for decisions over money today and money in the future together, they identified differential activity, *inter alia*, in the lateral prefontal cortex, which is one of the regions of the brain classically associated with "executive functions" involving trade-offs between competing goals.

So far so good, but how far have we come? It is premature to then conclude (p. 506) that "In economics, intertemporal choice has long been recognize as a domain in which 'the passions' can have large sway in

[14] Indeed, one can imagine standard experimental designs that can tease these apart, with uncertainty, lack of credibility, and transactions costs being applied differentially to the earlier option when $d > 0$. Implementing this thought experiment is not easy, however, since one has to convince subjects that there is greater uncertainty in the near future than the more distant future, and that is an unusual notion.

[15] They also report (p. 507, fn. 28) differential activation of the dorsal hippocampus. They reject this region for statistical reasons that are not clear, but this does not affect their main conclusion. The reason that this region is of some interest is that there is some evidence from rats suggesting that it might be used to *temporarily* store information used in spatial discrimination tasks (see White and Gaskin 2006).

affecting our choices (cites omitted). Our findings lend support to this intuition." Quite apart from the premise not being ascribed to by all economists, as distinct from those that push one version of what is happening with intertemporal choice behaviour, the findings do not speak to "passions" at all, or at least not obviously. The reward regions of the brain reflect the fact that money today will let me buy a sandwich today if I want, or something else as visceral as you like. And money today is surely more credible than money in the future, quite apart from any possible role of the executive function of the brain in translating it into a present value of money today. So the results are equally consistent with the view of hyerbolicky discounting as an artefact of asking subjects to compare "*m* good apples today" with "*M* poorer apples tomorrow", where apples are good or poor in terms of the credibility of me getting to eat them.

McClure *et al*. (2004: 507, fn. 29) make a show of addressing this issue, and ruling out the possibility that the credibility of immediate reward dominates choice:

> One possible explanation for increased activity associated with choice sets that contain immediate rewards is that the discounted value for these choice sets is higher than the discounted value of choice sets that contain only delayed rewards.

This could arise if the former choice sets were simply more salient, *ceteris paribus* the nominal amount of money, because they were more credible to the subject (or, equivalently, not subject to the subjective transactions costs of receiving them). But the manner in which this possibility is addressed is opaque at best:

> To rule out this possibility, we estimated discounted value for each choice as the maximum discounted value among the two options. We made the simplifying assumption that subjects maintain a constant weekly discount rate and estimated this value based on expressed preferences (best-fitting value was 7.5 % discount rate per week). We then regressed out effects of value from our data with two separate mechanisms. [. . .] Both of these procedures indicate that value has minimal effects on our results, with all areas of activation remaining significant . . .

The procedures in question, and omitted here, are alternative ways of including this estimate of "value" in the statistical model explaining differential voxel activation. There are many unclear aspects of this test. It appears to be making the maintained assumption that the subjects have exponential discounting functions, but attach a higher discount rate to the sooner options simply because they have high discount rates. So the implicit idea is that unless subjective value gets above some threshold, the reward regions of the brain will not differentially fire, and that the only reason they did fire with immediate payments is because of the higher

subjective value. But then why assume that every subject had the same discount rate of 7.5 % per week?

2.3 Other applications

Many applications of neuroeconomics follow the same pattern as the discounting application just considered. The use of neural data does not provide any insight in relation to the hypotheses being proposed, but is used to promote one favoured explanation even if it does not provide evidence that favours it in relation to known alternative explanations. There are some important exceptions to this pattern, where neural data is examined when subjects are "in equilibrium" as defined by some theory and "out of equilibrium", to detect possible differences in the processes at work and thereby generate testable hypotheses (e.g. Bhatt and Camerer 2005; Grether *et al*. 2007).

The Trust Game

The Trust Game offers a witches' brew of confounds. One player receives money from the experimenter (e.g. $10). He can then transfer some or all of that to another player. The experimenter then adds in some money to the transferred pie, typically scaling up the transferred amount by a factor of 3. Then the recipient decides what fraction of the enhanced pie, if any, to send back to the first player. The initial transfer is labelled as "trust", and the transfer back is labelled "trustworthiness", with the same skill that phrenologists of old used to label different lumps in the skull.[16] Of course, several things could motivate someone to send money, or send money back, and these are well known. Either player might be altruistic towards the other player. Or they might be spiteful towards the experimenter, wanting to extract more from him. Or they might be risk loving in the case of "trusting behaviour". Or they might view the one-shot game through the lens of a repeated game. The implication is that unless one wants to use the words "trust" and "trustworthiness" in the compounded sense of an amalgam of all of these motivations, one needs to design experiments to control for some or all of these other possible confounds, as in the designs of Cox (2004). Any trust experiment that does not do so is just adding to the pile of confusion. Does neuroeconomics improve on things?

[16] The concern here is not that there are confounds in the explanation of behaviour in the trust game. The task itself is rich in the sense that it identifies important characteristics of principal-agent problems in a crisp manner. Furthermore, any interesting institution or task will likely involve confounds at some level of analysis (e.g. the venerable double-auction, the work horse of the earliest writings in experimental economics, which still defies general behavioural formalization). The point is that we should not confuse labelling with explanation, as behaviourists do. Rubinstein (2006) makes the same point.

Kosfeld *et al*. (2005) and Zak *et al*. (2005) administer oxytocin to some subjects to see if it enhances trust and trustworthiness, respectively. Oxytocin is a neurally active hormone associated with bonding and social recognition. They find evidence that dosed subjects do send more in the Trust game, and that is not associated with them sending more in a nice control in which there is only risk involved. So the risk confound is controlled for, in aggregate "representative agent" terms.[17] They also find evidence that dosed subjects send more back in the Trust game, and here there is no need for a control for risk. But what about the other confounds? To take the obvious one, altruism, Zak *et al*. (2007) subsequently demonstrated that oxytocin had no effect on altruism, as measured in a simple Dictator game in which there is no strategic response involved, even if there is a social relationship.[18] So we later learn that it is probably not altruism, at least in terms of the representative agent. But there are no controls for other known confounds.

DeQuervain *et al*. (2004) also examine the Trust game, and trumpet the discovery of a "taste for punishment of unfair behaviour". They focus on the striatum, which we can stipulate for present purposes as being closely correlated with rewards in the brain. They see the striatum light up when subject A punishes subject B in a way that hurts B's payoffs but costs A nothing, and compare it to how the striatum lights up when the punishment is symbolic, and does not actually hurt B or cost A anything. They find that it lights up more in the initial condition than the latter. So what? If subjects viewed the game as repeated, then it could be a rational strategy to punish those that defect from a profitable strategy (Samuelson 2005). The underlying game was against different opponents in each round, so the subjects *should* have taken this into account and realized that it was not a repeated game. But many experimental economists, perhaps most notably Binmore (2007b: 1–22), would argue that we cannot easily displace field-hardened heuristics for playing repeated games when we drop subjects into an artefactual lab game. These are important methodological issues to resolve before we start adding (costly) brain scans to the analysis, since they perfectly confound inference. In effect, let's get an answer to this exercise in Binmore's (2007a) text on game theory before we start adding neural correlates:

> In laboratory studies, real people don't play the subgame-perfect equilibrium [in a game strategically similar to the Trust game]. The Humean explanation

[17] The risk control is only evaluated on an aggregate basis. It is possible that risk attitudes could explain the behaviour of some subjects at an individual level, and other factors explain the behaviour of other subjects. Thus the analysis is just incomplete, and not wrong as far as it goes.

[18] In this task one player sends money to the other player, and that is it. The other player does not get to reject it, so there is no strategic interaction.

> is that people are habituated to playing the fair equilibrium in repeated versions of the game. [. . .] comment on how people would use the words fairness, reputation, and reciprocity if the Humean explanation were correct. Why would this explanation be difficult to distinguish from the claim that people have a taste for a good reputation, fairness, or reciprocity built into their utility functions? (p. 349/350).

This is not to say that one or other of these competing explanations is better, or that indeed we should insist on just one explanation for these data. But it is clear that we have conceptual work to do before we fire up the scanner.

Game theory and the theory of mind

Rilling *et al*. (2004) scanned subjects playing the Ultimatum Bargaining game as well as the Prisoner's Dilemma. In each case the subject played one-shot games against a range of opponents. Some were supposedly human opponents, and a photo shown of the opponent, but the subjects were in fact deceived and the responses generated by computer. The responses were actually drawn from a distribution that reflected true human responses "in a typical uncontrolled version of the game, in which actual human partners" (p. 1696) made decisions. The scanned subjects also played against the computer, and were honestly told that.

Rilling *et al*. (2004) find that areas of the brain associated with a "theory of mind" differentially light up as the scanned subject receives feedback about the choices of the other player.[19] These areas received positive activation even when computers were generating the responses, which might seem odd. But consider the actual design, where actual human offers were fed to the scanned subjects by the computer: why would subjects not process them in the same way as they would if they had been generated there and then by humans, rather than by some distant human? Rilling *et al*. (2004) report *greater* activation in the deceptive treatment in which the scanned subjects were led to believe the responses came from actual humans, but they note (p. 1701) several plausible confounds: greater cognitive engagement when told they have a human opponent, which would also explain the activation; changing photos from round to round in the "human" treatments compared to the computer treatments, which instead had the same photograph of a boring computer instead.

[19] McCabe *et al*. (2001) conducted a similar exercise, but did not scan the subjects at the point where the feedback of the other player's choices was received. They detected differential activity in one of the regions associated with the "theory of mind" activity (the anterior paracingulate cortex), but not in two other regions noted by Gallagher and Frith (2003) in their review of the neuroscience literature.

3. THE TROUBLE WITH THOUGHT EXPERIMENTS

Gul and Pesendorfer (2008) stirred a debate, collected in Caplin and Schotter (2008), on the methodological status of neuroeconomics as positive economics.[20] They argued that positive economics is *defined* over choices, and makes no assumptions or predictions about the physiology or processes of the brain. Hence research generating *any* data on the brain, other than the choices that thinking agents make, is observationally irrelevant to the domain of economics. It is irrelevant, they claim, in the strong sense that it cannot be called on to support *or* refute economic theories about choice behaviour. Their argument deserves attention, because it quickly brings us to a better understanding of the potential role of neural data, as part of a broader drive to end the unproductive separation between theory and empirics in economics. The separation is not just a matter of specialization by comparative advantage, but has become embedded in the rhetoric used by economists.

3.1 Revealed preference and naked emperors

The beginning of the argument is familiar, from our earlier discussion of what utility functions are and what they are not (§ 1.2), and there is nothing to quarrel with here. But then they move beyond theory and implicitly consider the role of revealed preference as an empirical strategy for recovering preferences from observed choices in order to explain observed behaviour and, presumably, test theory:

> In the standard approach, the term utility maximization and choice are synonymous. A utility function is always an ordinal index that describes how the individual ranks various outcomes and how he behaves (chooses) given his constraints (available options). The relevant data are revealed preference data; that is, consumption choices given the individual's constraints. These data are used to calibrate the model (i.e., to identify the particular parameters) and the resulting calibrated models are used to predict future choices and perhaps equilibrium variables such as prices. Hence, standard (positive) theory identifies choice parameters from past behavior and relates these parameters to future behavior and equilibrium variables.

[20] Their target was wide-ranging. They also lumped in behavioural economics with neuroconomics, but there is a useful distinction between the two even if the *dramatis personæ* and marketing *modus operandi* are often the same. In addition, they considered the methodological status of neuroeconomics and behavioural economics as *normative* economics, but that involves separate issues. Finally, they point out how sloppy some of the economics has been in the neuroeconomics field, repeatedly setting up straw men to knock down, and those points are generally well taken.

One has the distinct feeling, however, that this is advice coming from non-practitioners.[21]

The problem comes when we try to make the theory operationally meaningful, in the sense of writing out explicit instructions on testing it, and the conditions under which it might ever be refuted. The separation between theory and empiricism cannot just be replaced by casual references to "revealed preference" or, worse, treated with an appended error term. This division of intellectual labor in economics, between theory and empirics, has become a serious hindrance to doing good economics.

Take revealed preference, a beautiful and powerful idea to help sort out misconceptions of behaviour, as illustrated in § 1.2, but practically useless as an empirical tool. First, it imposes very few constraints on behaviour without auxiliary assumptions about the domain of choice. Varian (1988) shows that if one only observes choices of a subset of goods, when the individual is actually making choices over a larger set, then revealed preference places essentially no restrictions on behaviour over the observed subset. Second, we have virtually no systematic theory of how to relate errors in the implications of revealed preference to a degree

[21] This is apparent when one comes across perfunctory, "motivating" references to the role of experiments in studies by theorists. For example, Gul and Pesendorfer (2006) "... develop and analyze a model of random choice and random expected utility. Modeling choice behavior as stochastic is a useful and often necessary device in the econometric analysis of demand. The choice behavior of a group of individuals with identical characteristics, each facing the same decision problem, presents the observer with a frequency distribution over outcomes. *Typically, such data are interpreted as the outcome of independent random choice by a group of identical individuals*. Even when *repeated decisions of a single individual are observed, choice behavior may exhibit variation and therefore suggest random choice.*" (p. 121; italics added). The first italicized empirical claim is nonsense, as a modicum of attention to the experimental literature would reveal: see Harrison and Rutström (2008) for a detailed review. The second italicized empirical claim completely ignores a large experimental literature, starting with the classic by Becker *et al*. (1963) and surveyed well by Hey (2005) and Loomes (2005). This active disinterest in literature slightly outside one's immediate formal interest leads to statements that leave the clear impression that the authors believe that one simple example is representative of the field: "Studies that investigate the empirical validity of expected utility theory predominantly use a random choice setting. For example, Kahneman and Tversky (1979) describe studies that report frequency distributions of choices among lotteries. These studies test expected utility theory by checking if the choice frequencies remain unchanged when each alternative is combined with some fixed lottery; that is, by testing our linearity axiom. Our theorems identify all of the implications of random expected utility maximization that are relevant for the typical experimental setting." As *any* student of experimental economics knows, tests of expected utility theory have been significantly more advanced than the example provided here (e.g. the fine review by Starmer 2000). It is hard to build a bridge between theory and evidence when theorists obviously have no real idea how insular they are.

of belief in the validity of the underlying theory.[22] Varian (1985) is the only attempt to address this matter, noting (p. 445) that in revealed preference tests the "... data are assumed to be observed without error, so that the tests are 'all or nothing': either the data satisfy the optimization hypothesis or they don't". The methods proposed, finding the smallest variations in the observed data to account for violations, is explicitly *ad hoc*: do you vary income levels, relative prices, quantities chosen, or some (weighted?) combinations of these?

Estimation of parameters is also something that involves a lot more than "calibration" using more and more refinements of revealed preference bounds. Many theorists are seduced by the attractive bounding of indifference curves that we present to our undergraduates, and think this ends up as an effective substitute for explicit econometric methods. Take the example that Gul and Pesendorfer (2008) offer:

> The standard approach provides no methods for utilizing nonchoice data to calibrate preference parameters. The individual's coefficient of risk aversion, for example, cannot be identified through a physiological examination; it can only be revealed through choice behavior. If an economist proposes a new theory based on nonchoice evidence then either the new theory leads to novel behavioral predictions, in which case it can be tested with revealed preference evidence, or it does not, in which case the modification is vacuous. In standard economics, the testable implications of a theory are its content; once they are identified, the nonchoice evidence that motivated a novel theory becomes irrelevant.

Consider, then, the estimation of risk aversion under standard Expected Utility Theory, essentially following the standard, pioneering approach of Camerer and Ho (1994) and Hey and Orme (1994).[23]

Assume some parametric utility function in which risk attitudes are determined by one parameter, such as a power function. Conditional on values for this parameter, Expected Utility Theory predicts that one lottery will be selected over the other, or that the subject will be indifferent (so nothing is predicted). Define a latent index of the difference in expected utility, ∇EU, favouring the lottery on the right (R) hand side of a choice. This latent index, based on latent preferences defined by the risk aversion parameter, is then commonly linked to the observed choices using a standard cumulative normal distribution function $\Phi(\nabla EU)$. This "probit" function takes any argument between $\pm\infty$ and transforms it into a number

[22] This is distinct from the power of tests of revealed preference, which will vary from instance to instance, as illustrated by Bronars (1987). That variation is not unusual, of course, and the reason one should do power calculations in general.

[23] Harrison and Rutström (2008; §2, §3) review elicitation procedures and statistical procedures for recovering estimates of risk attitudes in detail.

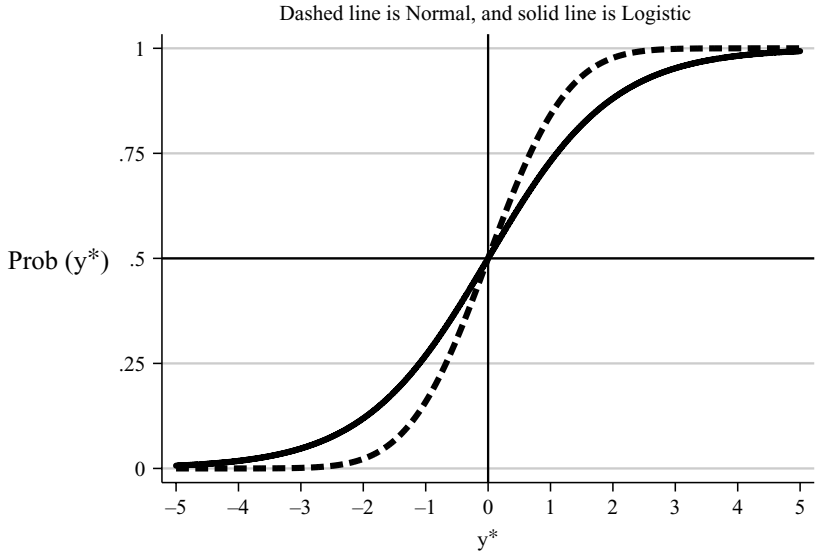

FIGURE 1. Normal and logistic cumulative density functions.

between 0 and 1 using the function shown in Figure 1. Thus we have the probit link function,

(1) prob(choose lottery R) = $\Phi(\nabla EU)$

The logistic function is very similar, as illustrated in Figure 1, and leads instead to the "logit" specification.

Even though Figure 1 is common in econometrics texts, it is worth noting explicitly and understanding. It forms the critical statistical link between observed binary choices, the latent structure generating the index y^*, and the probability of that index y^* being observed. Thus the likelihood of the observed responses, conditional on the validity of the Expected Utility Theory model and the specific utility parameterizations being true, depends on the estimates of the risk aversion parameter given this statistical specification and the observed choices. The "statistical specification" here includes assuming some functional form for the cumulative density function, such as one of the two shown in Figure 1. Formal maximum likelihood methods for estimating the preference parameter are reviewed in Harrison and Rutström (2008), and are not important here.

An important and popular extension of the core model is to allow for subjects to make some errors. The general notion of error is one that has already been encountered in the form of the statistical assumption that the probability of choosing a lottery is not 1 when the EU of that lottery exceeds the EU of the other lottery. This assumption is clear in the use of a link

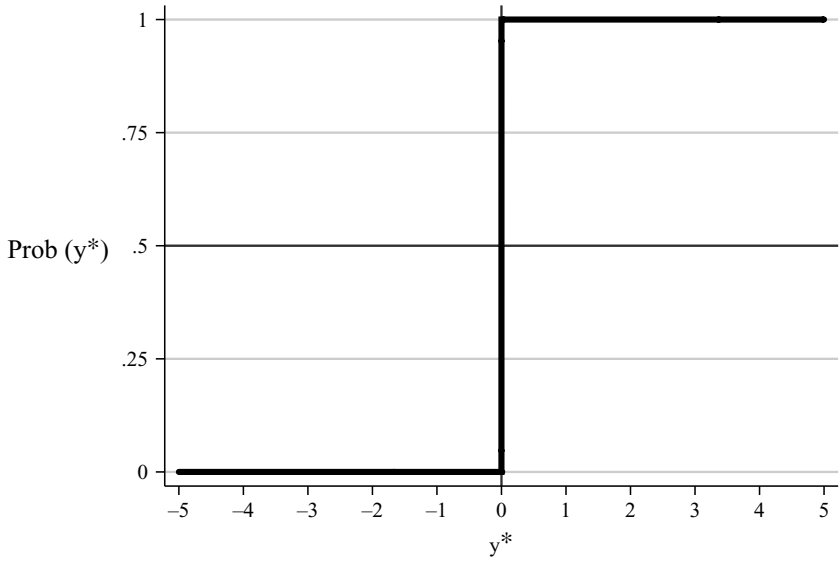

FIGURE 2. Hardnose theorist cumulative density function.

function between the latent index ∇EU and the probability of picking one or other lottery. If the subject exhibited no errors from the perspective of Expected Utility Theory, this function would be the step function shown in Figure 2: zero for all values of $y^* < 0$, anywhere between 0 and 1 for $y^* = 0$, and 1 for all values of $y^* > 0$. In contrast to (1), we then have the connection between preferences and data that a Hardnose Theorist would endorse:

(1a′) prob(choose lottery R) = 0 if ∇EU < 0

(1b′) prob(choose lottery R) = [0, 1] if ∇EU = 0

(1c′) prob(choose lottery R)=1 if ∇EU > 0

By varying the shape of the link function in Figure 1, one can informally imagine subjects that are more sensitive to a given difference in the index ∇EU and subjects that are not so sensitive. This is what a structural error parameter allows.[24]

[24] There are several species of "errors" in use (Loomes and Sugden 1998). Some place the error at the final choice between one lottery or the other after the subject has decided deterministically which one has the higher expected utility; some place the error earlier, on the comparison of preferences leading to the choice; and some place the error even earlier, on the determination of the expected utility of each lottery. The same ideas have

3.2 Hardnose theorists or mere theorists?

The problem with the cumulative density function of the Hardnose Theorist is immediate: it predicts with probability one or zero. The likelihood approach asks the model to state the probability of observing the actual choice, conditional on some trial values of the parameters of the theory. Maximum likelihood then locates those parameters that generate the highest probability of observing the data. For binary choice tasks, and independent observations, we know that the likelihood of the sample is just the product of the likelihood of each choice conditional on the model and the parameters assumed, and that the likelihood of each choice is just the probability of that choice. So if we have any choice that has zero probability, and it might be literally 1-in-a-million choices, the likelihood for that observation is not defined. Even if we set the probability of the choice to some arbitrarily small, positive value, the log-likelihood zooms off to minus infinity. We can reject the theory without even firing up any statistical package.

Of course, this implication is true for any theory that predicts deterministically, including Expected Utility Theory. This is why one needs some formal statement about how the deterministic prediction of the theory translates into a probability of observing one choice or the other, and then perhaps also some formal statement about the role that structural errors might play. In short, one cannot divorce the job of the theorist from the job of the econometrician, and some assumption about the process linking latent preferences and observed choices is needed. That assumption might be about the mathematical form of the link, as in (1), and does not need to be built from the neuron up, but it cannot be avoided. Even the very definition of risk aversion needs to be specified using stochastic terms unless we are to impose absurd economic properties on estimates (Wilcox 2008a, 2008b).

However, since one has to make some such assumption, why not see if insights can be gained from neuroscience? To some extent the emerging neuroscientific literature on the effect of reward systems on differential brain activition is relevant, since it promises to provide data that can allow us to decide between alternative propensities to make choices. Thus one subject, in one domain, might behave in the manner of Figure 2, one subject more in the manner of one of the curves in Figure 1, and another subject more in the manner of the other curve in Figure 1. With structural noise parameters, and more flexible specifications of the cumulative density

been applied in game theory in the form of "quantal response equilibria". In that context there is also a lively debate over the extent to which one can separate assumptions about stochastic specifications from substantive predictions of the core theory: contrast Haile, Hortaçsu and Kosenok (2008) and Goeree, Holt and Palfrey (2005).

function, one could allow a myriad of linking functions to be data driven. Nobody is saying that neural data is sufficient to do this now, and I can think of more efficient ways to go about estimating it before I would put a subject in a scanner, but the methodological point is to identify what role neural data *could* play. The general point is to see thought experiments for what they are, and to see why we do experiments with real subjects.

3.3 The limits of thought experiments

Sorenson (1992) presents an elaborate defense of the notion that a thought experiment is really just an experiment "that purports to achieve its aim without the benefit of execution" (p. 205), and that such experiments can be viewed as "slimmed-down experiments – ones that are all talk and no action". This lack of execution leads to some practical differences, such as the absence of any need to worry about luck affecting outcomes.[25] A related trepidation with treating a thought experiment as just a slimmed-down experiment is that it is untethered by the reality of "proof by data" at the end. But this has more to do with the aims and rhetorical goals of doing experiments. As Sorenson (1992: 205) notes:

> The *aim* of any experiment is to answer or raise its question rationally. As stressed [earlier...], the *motives* of an experiment are multifarious. One can experiment in order to teach a new technique, to test new laboratory equipment, or to work out a grudge against white rats. (The principal architect of modern quantum electrodynamics, Richard Feynman, once demonstrated that the bladder does not require gravity by standing on his head and urinating.) The distinction between aim and motive applies to thought experiments as well. When I say that an experiment 'purports' to achieve its aim without execution, I mean that the experimental design is presented in a certain way to the audience. The audience is being invited to believe that contemplation of the design justifies an answer to the question or (more rarely) justifiably raises its question.

In effect, then, it is *caveat emptor* with thought experiments – but the same homily surely applies to any experiment, even if executed.

[25] Another difference, noted by Harrison and List (2004; § 8), is that thought experiments actually require *more* discipline if they are to be valid. In his Nobel Prize lecture, Smith (2003; p.465) notes that "Doing experimental economics has changed the way I think about economics. There are many reasons for this, but one of the most prominent is that designing and conducting experiments forces you to think through the process rules and procedures of an institution. Few, like Einstein, can perform detailed and imaginative mental experiments. Most of us need the challenge of real experiments to discipline our thinking." There are, of course, other differences between the way that thought experiments and actual experiments are conducted and presented. But these likely have more to do with the culture of particular scholarly groups than anything intrinsic to each type of experiment.

This might all seem like an exceedingly fine point until we consider the link between theory and evidence. We rejoice in an intellectual division of labour between theorists and applied economists, but the absence of explicit econometric instructions on how to test theory has led to some embarrassing debates in economics.[26] To avoid product liability litigation, it is standard practice to sell commodities with clear warnings about dangerous use, and operating instructions designed to help one get the most out of the product. Unfortunately, the same is not true of economic theories. When theorists undertake thought experiments about individual or market behaviour they are positing "what if" scenarios which need not be tethered to reality. Sometimes theorists constrain their propositions by the requirement that they be "operationally meaningful", which only requires that they be *capable* of being refuted, and not that anyone has the technology or budget to actually do so.

4. ECONOMIC BEHAVIOUR AS ALGORITHMIC PROCESS

If the current promotional material and substantive examplars of neuroeconomics leave us disappointed, should we conclude that neuroeconomics is unlikely to *ever* contribute anything valuable? Smith (2007: 313) draws the conclusion that

> Neuroeconomics will not achieve distinction in a focus confined to correcting the 'errors' believed to pervade professional economics of the past, an exercise of interest to a narrow few. Nor will notoriety likely stem from better answers to the traditional questions. Rather, neuroeconomic achievement more likely will be determined by its ability to bring a new perspective and understanding to the examination of important economic questions that have been intractable for, or beyond the reach of, traditional economics. Initially new tools tend to be applied to the old questions, but [. . .] their

[26] And not just limited to experimental economics. In a famous study of tests of the restrictions of optimization in demand systems, Deaton (1974) reports a rejection of homogeneity. The lack of methodological direction is then palpable in these helpless complaints: "Homogeneity is a very weak condition. It is essentially a function of the budget constraint rather than the utility theory and it is difficult to imagine *any* demand theory which would not involve this assumption. Indeed, to the extent that rationality has any place in demand analysis, it would seem to be contradicted by non-homogeneity. [. . .] Now we may accept this rejection, implying our acceptance of the framework within which the experiment was carried out, or we may refuse to do so, claiming that the experiment was wrongly performed and that a correct experiment would have led to the opposite result. The first implies the acceptance of non-homogeneous behavior and would seem to require some hypothesis of 'irrational' behavior; this is not an attractive alternative. We are thus left to excuse our failure but without further information it is difficult to do this in a convincing fashion. [. . .] Whether or not we are justified, this is what we shall do here. The formal rejection must go on record but it would seem that to continue with further tests having imposed homogeneity is more acceptable than turning away together." (p. 362/3) Tiger Woods just topped his drive.

> ultimate importance emerges when the tools change how people think about their subject matter, ask new questions, and pursue answers that would not have been feasible before the innovation. Neuroeconomics has this enormous nonstandard potential, but it is far too soon to judge how effective it will be in creating new pathways of comprehension.

The time is nigh to start being explicit about these "new questions," and to think about what form these "previously infeasible answers" will take. The approach proposed here is to to *formally view economic behavior as the outcome of algorithmic processes*, and is consistent with many of the intellectual paths leading to neuroeconomics. The idea is familiar to many behavioural economists already, but only as metaphor.[27]

4.1 Behaviour as the outcome of one or more processes

How do individuals make good decisions, and why do they also make bad decisions? Standard economic theory has a relatively simple hypothesis that generates answers to the first question up to a point, but that is silent with respect to the second question. Since experimental data *appears* to provide a wealth of observations of each type of decision, a good deal of frustration has swirled around the effort to interpret experimental data in terms of standard economic theory. Indeed, it is fair to say that experimental data has provided the lightning rod for debates over the behavioural relevance of standard economic theory.[28]

The general answer offered by economic theory to the first question is that individuals act *as if* they *have* optimi*zed* a well-behaved objective function subject to some well-defined constraints. This general hypothesis, along with some further structure on the nature of the objective functions and/or the constraints, provides testable restrictions on observable behaviour.[29] Moreover, there is no presumption that there is only one formulation of the objective function and constraints that can account for the observed behaviour.

Whenever there appears to be evidence that individuals make bad decisions, the data tend to be interpreted in one of two ways. Either the data

[27] The formal use of algorithmic machines, automata, has provided many of the major insights in repeated game theory: Binmore (2007a: 328ff.).

[28] Many of these debates have not been joined in any productive sense, so I do not want to endorse them as being particularly useful. In fact, some, such as the debates on discounting behaviour, risk aversion "calibration", and the existence of loss aversion, have taken on a resolutely thuggish tone that discourages useful discussion.

[29] Sometimes it simply takes a long time to work out the testable implications of even classic problems in all of their generality. In the meantime we risk rejecting or accepting theory based on incomplete tests. For example, Paris and Caputo (2002) provide a general characterization of a problem first posed by Samuelson and Patinkin in the late 1940s. Many other examples derive from the development or correct application of econometric tools well after the initial debates had dulled.

are questioned as being a valid test of the theory, or the theory is discarded. These are extreme positions, and most serious researchers appropriately qualify their views with the usual double negatives, but in general terms this seems to be the way the empirical tension is resolved.

One possible alternative is to relax the perspective that economic theory has on the behavioural process being observed, and to interpret behaviour *as if* the subjects *are* optimiz*ing* a well-behaved objective function subject to some well-defined constraints. Keep the "as if" preface, but just shift the tense slightly so that we *view the observed subject behavior as potentially being iterations of some algorithmic process rather than as the end-point of that process.* An alternative way of viewing this slight shift is to think of the researcher as exploiting the way that the researcher might himself solve the problems that the subject is *modelled* as solving so as to gain insights into how the *subject* might be solving the problem. The obvious methodological parallel is to the "rational expectations" insight into modelling the beliefs of subjects in a system.

This emphasis on behaviour as reflecting an algorithmic process is one that is consistent with the rhetoric of many neuroeconomists. One difference is to take the idea of algorithms, and their epistemologically poor cousins, heuristics, seriously as a guiding framework for analysis. Thus one thinks of algorithms as more than metaphors, and exploits the fact that we know a lot about when algorithms work well and when they do not.

4.2 Algorithms versus heuristics?

If one adopts an algorithmic perspective on the decision-making process, it does not follow that one must rule out attention to heuristics. The key distinction between an algorithm and a heuristic has to do with the knowledge claim that they each allow one to make. If an algorithm has been applied correctly, then the result will be a solution that we know something about. For example, we may know that it is a local optimum, even if we do not know that it is a global optimum. Heuristics are lesser epistemological beasts: the solution provided by a heuristic has no claim to be a valid solution in the sense of meeting some criteria. In the computational literature, if not some parts of the psychological literature, heuristics are akin to "rules of thumb" that simply have good or bad track records for certain classes of problems.[30] The track record may be defined

[30] The psychological literature on the heuristics is divided into two major, warring camps. One, which behavioural economists represent as if it is the settled consensus of psychology, is of course associated with Kahneman and Tversky (1996), and is known as the "heuristics and biases" research programme. The other is known as the "fast and frugal heuristics" research programme, and is associated with Gigerenzer (1996).

in terms of the speed of arriving at a candidate solution, or the ease of application.

The line between the two is not always clear. Many algorithms can be heuristically applied. For example, one of the most popular ways to start solving an integer programming problem is to define a "relaxed" version of the problem that does not constrain the solution variables to take on discrete values, solve it using some appropriate algorithm for the relaxed problem, and then do a systematic local search for discrete solutions in the neighbourhood of the solution to the relaxed problem. If the result of this procedure is a proposed solution that is not in fact a global optimum, then one should not blame that on the application of the algorithm to the relaxed problem (assuming it was implemented correctly).

Similarly, most of the algorithms in widespread use to solve real problems utilize one or more heuristics to guide their behaviour "under the hood." Most of the solvers for the GAMS software package, for example, are robust precisely because they have built in much of the experience gleaned by their authors from solving a wide array of instantiations of their problem class. The efficient choice of "stepping size", for example, can be critical to the speedy application of algorithms for non-linear programming problems. In some difficult problems these algorithms spend a great deal of time explicitly "back-tracking" when iteration steps do not produce the desired improvement in the objective function; but most of the time simple adjustment rules suffice to detect multidimensional escapes from (very) local plateaus. Above all, differences in the knowledge claims that results from applying heuristics or algorithms should not divert from the essential complementarity of the two.

4.3 Homotopies and path-following

The homotopy approach to solving non-linear systems of equations provides a powerful generalization of many existing solution methods, and a general framework to see the complementary roles of economics and several other sciences.[31] To fix ideas, consider a system that has n variables and n equations, and let $x = (x_1, x_2, \ldots, x_n)$ define the values of the variables. We seek the solution values of x, denoted x^*, that solve the $n \times n$ system of non-linear equations $F(x) = 0$.

Let $H(x(t), t)$: $\mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$ define a homotopy function over x, where t is a scalar parameter that determines the path to be followed from some initial solution at $t = 0$ to the final solution at $t = 1$. Specifically, we define the homotopy function so that at $x = x^0$ the original system $F(x)$ has a "simple solution". Define this system by $E(x) = 0$, where we know that x^0 solves it, so that $H(x(t), 0) = H(x^0, 0) = E(x)$. We examine below what we

[31] Also known as continuation methods, they are well-known to economists in numerical work: see Judd (1998; p.176ff.).

mean by a simple solution. We also want the homotopy function to solve the original system $F(x) = 0$ when $t = 1$, so that $H(x(t), 1) = H(x^*, 1) = F(x)$. And along the path defined by tracing out values of t between 0 and 1, we want the homotopy function to define the deformed system obtained by "taking *some* weighted average" of E(x) and F(x), so that $H(x(t), t) = 0$ for $0<t<1$ as well. The idea of taking a weighted average of the initial solution and the final solution is only a metaphor, but a useful one. When we specify particular homotopy functions in examples below it will be clear in what respects it is a metaphor and in what respects it is not.

The requirements for the homotopy function are (i) that H(x, 0) be easily solved for x^0; (ii) that the solution to $H(x, 1) = 0$, x^*, is the solution to $F(x) = 0$; and (iii) that the path x(t) leads from x^0 to x^* as t increases. If we can find some representation of the original system $F(x) = 0$ that is easy to solve, and if we can define a path from that initial representation to the original system that is well-behaved in the sense that we can easily follow it, then we will have a way of finding the solution to the original system.

What do we mean by a simple solution to the initial, deformed system E(x)? Literally, any solution that the human machine, aided with field referents, firing executive-region neurons, and even a dose of cognitive serendipity, can ascertain. To take some concrete and familiar examples (Garcia and Zangwill 1981: ch.1), the Newton homotopy lets one pick an arbitrary x^0, providing it has the same dimensionality as the final solution, and then start from it. Given x^0, you calculate $F(x^0)$, adopt $E(x) \equiv F(x) - F(x^0)$, and the homotopy function then becomes $H(x, t) = F(x) - (1-t) F(x^0)$. So this method would be an appropriate formalization if one had some reason to think that subjects had some focal point solution for their calculations: in a first-price sealed-bid auction over private values, for example, the bidder's private value itself is an obvious suggestion. An alternative homotopy, also attractive because it allows one to start at an arbitrary x^0, is the Fixed-Point homotopy. In this case let $E(x) \equiv x - x^0$, and use the function $H(x, t) = (1-t)(x - x^0) + tF(x)$. Or the final system F(x) might be seen, psychologically or mathematically, to be similar to some special-case with known solution E(x), and the Linear homotopy can be used in which $H(x, t) = tF(x) + (1-t)E(x) = E(x) + t[F(x) - E(x)]$. This might be appropriate if there was some special case of the final system that was relatively easy to solve, or familiar from field contexts, such as infinitely repeated versions of a one-shot game. Note well that in each example E(x) is a deformation and *representation* of F(x).

It is easy to see how this formalization neatly admits many of the interests of psychologists, behavioural economists, neuroeconomists and mainstream economists.

Psychologists have spent a lot of time studying task representation and recognition, which can be taken as the psychological counterpart to selecting the starting point of the homotopy. There are relatively few formal restrictions on what constitutes a good starting point: for instance,

Newton-type homotopies and Fixed-Point homotopies allow virtually arbitrary values for the solution variables to be used. So this is fertile ground for similarity relations to be applied to explain the process by which individuals adopt "deformed representations" of the original problem that can be solved quickly by the neural hardware and software that the brain provides. Sometimes this may be a conscious choice of a deformation, or it might be a representation that is dredged up by the brain from the evolutionary bog without any executive function operating at all. Or it might be conscious and deliberate for some individuals, or some task domains, and automatic for others. In any event, we have the beginnings of a process that can lead to considerable heterogeneity in the path to be followed.

Without naming it as such, experimental economists and cognitive psychologists have spent a lot of time studying when subjects apply more or less effort to solving problems. These measures of effort can be viewed as indicia of the fraction of the path followed and the effort assigned to coming up with the initial representation. In some cases the measures are as direct as response time (Luce 1986: Wilcox 1993; Rubinstein 2007) or mouse clicking (Johnson *et al*. 2002).

Insights about the type of path being followed also come from applied mathematics. Allgower and Georg (2003) draw a distinction between two types of ways in which one can follow a homotopy path. One is the class of "predictor-corrector" methods, which attempt to closely trace out some smooth path defined by the homotopy function. These methods generally rely on the ability of the system to undertake lots of very small steps if the path is non-linear, and presume smoothness of the path. On the other hand, one might use "piecewise linear" methods that derive from convenient triangulations of the space over which the path is defined. These triangulations have the advantage that they might reflect efficient ways to store information: all that is needed is the information on the current simplex, and the rules needed to go to the next one (the "pivoting step"). So they might reflect pre-existing neuronal hardware developed for some other purpose, reflecting the evolutionary nature of the human brain (Dehaene *et al*. 1999; Linden 2007).

Similarly, the issue of "stopping rules" has been extensively studied by decision theorists (von Winterfeldt and Edwards 1982, 1986) and experimental economists (Harrison 1989, 1992: Merlo and Schotter 1992). The major issue has been to identify the metric of evaluation that subjects employ when evaluating alternative off-equilibrium choices, but there also remain mundane algorithmic issues such as identifying solution tolerances.

Homotopy methods can be used to solve an extraordinarily wide range of problems of interest to economists. Constrained optimization problems are standard fare, but equilibrium problems encountered in

general equilibrium analysis and applied game theory can also be solved using path-following methods. Indeed, many existing solution methods are effectively path-following methods even if they have not traditionally been presented that way. For example, the "tracing procedure" of Harsanyi (1975) is readily seen to be a homotopy method. Garcia and Zangwill (1981: Part II) provide an extensive series of examples. Therefore, one advantage of the general path-following approach is that it offers a relatively unified approach to wide classes of problems in economics.

An explicitly algorithmic framework admits of the latent process assumed by economists as a special case, but allows space for insights from psychology, cognitive science, and, yes, even neuroscience.[32]

4.4 Hardware, software, and explanation

One feature of the performance of algorithms, of some significance for the manner in which neuroeconomics seeks to explain behaviour, is the relationship between hardware and software. Understanding the physics of computers does not help us much to understand what makes one algorithm better than another. But once we know what we want an algorithm to do, and the various ways it can achieve its goal, we do care about the hardware. We care about the speed of the processor, we care about the available of concurrent processors, we care about input-output speed and hard disk capacity, and we might even care about rendering speed for graphics. This is exactly the concern of Marr (1982: 27ff.):

> Trying to understand perception by studying only neurons is like trying to understand bird flight by understanding only feathers: It just cannot be done. In order to understand bird flight, we have to understand aerodynamics; only then do the structure of feathers and the different shapes of birds' wings make sense. More to the point, we cannot understand why retinal ganglion cells and lateral geniculate neurons have the receptive fields that they do just by studying their anatomy and physiology. We can understand how these cells

[32] Many economists have a pinched understanding of what cognitive psychology is all about, perhaps from relying on sources such as Rabin (1998). The introductory text by Anderson (2000), which I recommend to my experimental economics students, contains chapters full of important insights about the processes I have in mind: perception; attention and performance; perception-based knowledge representations; meaning-based knowledge representations; human memory: encoding and storage; human memory: retention and retrieval; problem solving; development of expertise; reasoning and decision-making; language structure; language comprehension; and individual differences in cognition. Smith (1991) provides an early attempt, still being pursued by Smith (2003, 2007), to reorient experimental economics towards a *deeper* connection to psychology: "Experimental economics can benefit greatly from the criticisms of psychologists, but in order for this to occur, their knowledge and understanding of the literature and its motivation will have to move beyond the superficial level of familiarity exhibited in [this symposium issue] [. . .] We need the help of psychologists, undeflected by battles with straw men." (p. 893/894).

> and neurons behave as they do by studying their wiring and interactions, but in order to understand *why* the receptive fields are as they are – why they are circularly symmetrical and why their excitatory and inhibitory regions have characteristic shapes and distributions – we have to know a little of the theory of differential operators, band-pass channels, and the mathematics of the uncertainty principle.

So look at brains, but understand why we do so. This perspective is a "top down" one from a functional perspective, not a reductionist "bottoms up" approach.[33] But it does not ignore or belittle the role of the hardware; instead, it just puts it in its place when it comes to trying to explain why behaviour occurs in the manner in which it does.

In fact, neuroeconomics should have been ideally positioned to pursue this approach, since economists have such a strong sense of what problems economic behaviour can be usefully viewed as solving. This is not to take a position on whether Expected Utility Theory is the best model of choice under uncertainty for every individual in every task domain, or whether the straw-man of self-regarding preferences is the best one to assume, to take two examples, but to point out that we already had a lot of that work done. This caveat is critical, since it is one of the later criticisms of the approach developed in Marr's name,[34] that it presumed the definition of "the" computational goal of the "agent". Economists have learned, from internal debates, how to extend their standard paradigm in astonishing ways (Stigler and Becker 1977), how to apply their own optimization

[33] Marr's work is reviewed in detail by Glimcher (2003: ch. 6), with a refreshing balance of respect for the intellectual milieu in which it was developed as well as the life-cycle of intellectual fads. Marr and Poggio (1976) was the first paper to propose this framework. Glimcher (2003: ch. 7) discusses one of the main weaknesses of Marr's approach for the purposes of biology: the need to delimit the computational scope of the objective function being studied, so that one could focus on a specific neurobiological module. In Marr and Poggio (1976) and Marr (1982) this problem did not arise, in part because they were just claiming that the module used in previous research on vision was too narrow, and they did not need to say precisely how broad it had to be to make their case. But it also failed to arise, as noted by Glimcher (2003: 143), because Marr looked at vision as a computer scientist rather than as a biologist, so the questions that he was trying to answer were conceptually different and did not need a physical, biological counterpart. Economists are much more like computer scientists in this respect than biologists, or should be. On the other hand, modern linguists are more like computer scientists than biologists, and have still devoted a lot of time to debating the utility of the notion of a "linguistic module" in the brain. But properly understood as metaphor, rather than biological prerequisite or even surgical marker, it has generated important behavioral and neural hypotheses: see Anderson (2000: ch. 11, 12).

[34] We must be careful here, since the history of economic thought teaches us that Keynes was no Keynesian, and even Marx is reported to have informed his own followers that "Ce qu'il y a de certain c'est que moi, je ne suis pas Marxiste." [If anything is certain, it is that I myself am not a Marxist] (http://www.marxists.org/archive/marx/works/1882/letters/82_11_02.htm).

paradigm to more elaborate objective functions and constraints (e.g. the alternatives to Expected Utility Theory), and even how to contemplate the possibility that there might be multiple computational goals at work (mixture specifications, as in Harrison and Rutström 2005).

4.5 Agency

One way to couch the debate over the role of neuroeconomics is to see it as a debate over the meaning of "agency" in economics: who is the economic agent? To many economists this seems like a non-question, but it is front and centre in philosophical debates over economics and cognitive science. Ross (2005) provides a detailed review of the issues, which run deep in philosophy.[35]

One perspective is to think of the brain as made up of multiple selves or agents, so that agency is defined in terms of the part of the brain that makes specific decisions.[36] Some use this as metaphor, and others see it as more literal. The concept of "dual selves" has a long lineage in behavioural economics, and findings from neuroscience certainly suggest that multiple brain systems interact when subjects make economic decisions (Cohen 2005). An alternative interpretation of the concept "dual selves" would be a single decision maker that has dual cognitive processes that are activated under different conditions. This interpretation is consistent with the literature on dual process theories of mind in psychology and

[35] My objective here is just to plant a flag on this issue: there is a large, difficult, tendentious and important literature here. As Judd (1998: 27) dryly warns, prior to the exercises at the end of the first chapter of his advanced text on numerical methods, "These first exercises are warmup problems; readers having any difficulty should stop and learn a computing language."

[36] Alternatively, one might posit agency as being defined over groups of individuals, focussing again on the initial homotopy representation of the decision process. An intriguing hypothesis to emerge from this perspective is that subjects might deliberately and consciously adopt a conventional representation of the formal game presented to them in order to make it easier to solve *and that they would choose a conventional representation that maximizes their joint payoff* if one existed. They are already assumed by the algorithmic approach to be coming up with some initial representation that is easy to solve from a computational perspective, so it is but a short step to assume that they might adopt a representation that also maximizes their expected joint payoff if they had a choice of alternative conventional representations. Such representations would, to be sure, entail a "meeting of the minds," but there is stunning evidence from Mehta, Starmer and Sugden (1994) that subjects can jointly identify salient focal points in simple coordination games. The rational deformation hypothesis seeks to put some structure on this choice behaviour, by hypothesizing that the subjects as a group behave as if they solve some initial coordination game defined over alternative representations of the experimental task. This hypothesis is consistent with the philosophical literature on frames for coordination games and the possibility of group agency as an explanation for rational cooperation (Sugden 2003; Bacharach 2006).

economics (e.g. see Lopes 1995; Barrett *et al*. 2004; Benhabib and Bisin 2005; and their references to the older literature). It also lends itself to formalization in certain structured settings (e.g. Andersen *et al*. 2008a) and more generally using mixture specifications defined over multiple latent decision-making processes (e.g. Harrison and Rutström 2005; Andersen *et al*. 2007).

4.6 Multiple levels of selection

A fundamental tenet of neuroeconomics appears to be the idea that it is neuronal activity that ultimately determines if we behave in one way or another in the economic domain. The algorithmic approach offers another model of selection of behaviour: in some cases the process is causal in the direction suggested by neuroeconomists, but in other cases the process might have the reverse causality. These processes might operate concurrently, sometimes over very short periods of time, and sometimes over extended evolutionary time (e.g., Keller 1999). Sunder (2006: 322) identified this as one of the insights from an algorithmic perspective on economic behaviour:

> The marriage of economics and computers led to a serendipitous discovery: there is no internal contradiction in suboptimal behaviour of individuals yielding aggregate-level outcomes derivable from assuming individual optimization. Individual behaviour and aggregate outcomes are related but distinct phenomena. Science does not require integration of adjacent disciplines into a single logical structure. As the early-twentieth-century unity of science movement discovered, if we insist on reducing all sciences to a single integrated structure, we may have no science at all. In Herbert Simon's (1996: 16) words: "This skyhook-skyscraper construction of science from the roof down to the yet unconstructed foundations was possible because the behavior of the system at each level depended on only a very approximate, simplified, abstracted characterization of the system at the level next beneath. This is lucky; else the safety of bridges and airplanes might depend on the correctness of the 'Eightfold Way' of looking at elementary particles." This is the story of how we found that economists can have their cake while psychologists eat it too.

And there might even be some cake left for neuroscientists: if not, blame (cognitive) psychologists!

5. CONCLUSIONS

From the perspective of economists, the neuroeconomics literature seems to have used a production function with a sub-optimal mix of human capital and physical capital, in a blushing fascination with the toys of neuroscience. The result has been dazzling images of light bulbs popping on in different parts of the brain, but unimpressive economics. Straw-men

are erected as null hypotheses, multiple alternative hypotheses are ignored and left behind as the literature cites itself in a spiral, and known confounds are glossed. As the behavioural economics literature demonstrated, however, we already knew how to do poor economics (and get it published). The fear is that the impressive and important machinery of neuroscience will make it even harder for anyone to know what passes for scientific knowledge in economics and what is just great story-telling.

We can put the academic marketing of neuroeconomics aside. It almost seems unfair to put some of those claims on display, but, like the assertions of the behaviourists, they have taken hold in many quarters as knowledge claims when they are just "chloroform in print". Since economists have important and serious questions to get on with, the opportunity cost of these diversions has just become too great to ignore.

The more important business is to decide what to make of the substantive claims of neuroeconomics. In this respect the evaluation is mixed.

As an economist I do not learn much on the broad substantive issues reviewed, and these cover what should be the low-hanging fruit for neuroeconomics. The lack of insight does not primarily come from unfamiliarity with the details of the methods: some of those details bother me, and need exposition by the economists on these teams, but that is not in the end decisive. My main concern is whether neuroeconomists have added insight to already-confused experimental designs, or just covered up those confusions and promoted one plausible story over another. I conclude the latter, unfortunately. Obviously there is no need to add neural correlates to pre-confused designs.

But the potential remains. I reject the view that neural data *must* be irrelevant to economics as needlessly isolationist. I do not take the free-disposability view that *any* data is useful data until proven otherwise, implying that we should just collect it anyway and decide later if it was useful; that is a poor model for advancement of study in any field.[37] Instead, I encourage a restatement of the formal processes by which agents make economic decisions, so that we can better see what questions neural data can provide insight into. This restatement does not mean rejecting what we have already in mainstream economics, but viewing it as a special case which may or may not be applicable in certain domains. The framework I have in mind leaves a clear role for neural data, but a more urgent role for proper, sustained communication between economics and cognitive psychology.

[37] There is also an opportunity cost of collecting these data. My concern is not so much with the "out of pocket" costs, which have been sizeable in the past but justifiable.

REFERENCES

Allgower, E. L. and K. Georg. 2003. *Introduction to numerical continuation methods*. Phildaelphia: Society for Industrial and Applied Mathematics.

Andersen, S., G. W. Harrison, M. I. Lau and E. E. Rutström. 2007. Behavioral econometrics for psychologists. Working Paper 07-04. Department of Economics, College of Business Administration, University of Central Florida; *Journal of Economic Psychology*, forthcoming.

Andersen, S., G. W. Harrison, M. I. Lau and E. E. Rutström. 2008a. Eliciting risk and time preferences. *Econometrica* 76: 583–618.

Andersen, S., G. W. Harrison, M. I. Lau and E. E. Rutström. 2008b. Lost in state space: are preferences stable? *International Economic Review* 49: 1091–1112.

Anderson, J. R. 2000. *Cognitive psychology and its implications*, 5th edn. New York: Worth.

Andreasen, N. C., S. Arndt, V. Swayze, T. Cizadlo, M. Flaum, D. O'Leary, J. C. Ehrhardt and W. T. C. Yuh. 1994. Thalamic abnormalities in schizophrenia visualized through magnetic resonance image averaging. *Science* 266: 294–8.

Bacharach, M. 2006. *Beyond individual choices: teams and frames in game theory*. Princeton, NJ: Princeton University Press.

Barrett, L. F., M. M.Tugade and R. W. Engle. 2004. Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin* 130: 553–73.

Benhabib, J. and A. Bisin. 2005. Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption-saving decisions. *Games and Economic Behavior* 52: 460–92.

Becker, G. M., M. H. DeGroot, H. Morris and J. Marschak. 1963. Stochastic models of choice behavior. *Behavioral Science* 8: 41–55.

Berridge, K. C. 1996. Food reward: brain substrates of wanting and liking. *Neuroscience and Biobehavioral Reviews* 20: 1–25.

Bhatt, M. and C. F. Camerer. 2005. Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. *Games and Economic Behavior* 52: 424–59.

Binmore, K. 2007a. *Playing for real. A text on game theory.* New York: Oxford University Press.

Binmore, K. 2007b. *Does game theory work? The bargaining challenge*. Cambridge, MA: MIT Press.

Bolton, P. and M. Dewatripont. 2005. *Contract Theory*. Cambridge, MA: MIT Press.

Bowman, F. D., B. Caffo, S. S. Bassett and C. Kilts. 2008. A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage* 39: 146–56.

Bronars, S. G. 1987. The power of nonparametric tests of preference maximization. *Econometrica* 55: 693–8.

Bullmore, E., M. Brammer, S. C. R. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard and P. Sham. 1995. Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine* 35: 261–77.

Camerer, C. F. and T.-H. Ho. 1994. Violations of the betweeness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty* 8: 167–96.

Camerer, C. 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Camerer, C. 2008. The case for mindful economics. In *Foundations of positive and normative economics*, ed. A. Caplin and A. Schotter. New York: Oxford University Press.

Camerer, C., G. Loewenstein and D. Prelec. 2004. Neuroeconomics: why economics needs brains. *Scandinavian Journal of Economics* 106: 555–79.

Camerer, C., G. Loewenstein and D. Prelec. 2005. Neuroeconomics: how neuroscience can inform economics. *Journal of Economic Literature* 43: 9–64.

Caplin, A. and A. Schotter, Andrew, eds. 2008. *Foundations of positive and normative economics*. New York: Oxford University Press.

Chambers, R. G. and J. Quiggin. 2000. *Uncertainty, production, choice, and agency: The state-contingent approach*. New York: Cambridge University Press.
Chandrasekhar, P. V. S., C. M. Capra, S. Moore, C. Noussair and G. Berns. 2008. Neurobiological regret and rejoice functions for aversive outcomes. *NeuroImage* 39: 1472–84.
Cohen, J. D. 2005. The vulcanization of the human brain: A neural perspective on interactions between cognition and emotion. *Journal of Economic Perspectives* 19(4): 3–24.
Coller, M. and M. B. Williams. 1999. Eliciting individual discount rates. *Experimental Economics* 2: 107–27.
Cox, J. C. 2004. How to identify trust and reciprocity. *Games and Economic Behavior* 46: 260–81.
Crum, W. R., L. D. Griffin, D. L. G. Hill and D. J. Hawkes. 2003. Zen and the art of medical image registration: correspondence, homology, and quality. *NeuroImage* 20: 1425–37.
Deaton, A. S. 1974. The analysis of consumer demand in the United Kingdom, 1900–1970. *Econometrica* 42: 341–67.
Dehaene, S. and J.-P. Changeux. 1997. A hierarchical neuronal network for planning behavior. *Proceedings of the National Academy of Sciences USA* 94: 13293–13298.
Dehaene, S., E. Spelke, P. Pinel, R. Stanescu and S. Tsivkin. 1999. Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science* 284: 970–974.
Delgado, M. R., L. E. Nystrom, C. Fissell, D. C. Noll and J. A. Fiez. 2000. Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of Neurophysiology* 84: 3072–7.
DeQuervain, D., U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck and E. Fehr. 2004. The neural basis of altruistic punishment. *Science* 305: 1254–8.
Dickhaut, J., K. McCabe, J. C. Nagode, A. Rustichini, K. Smith and J. V. Pardo. 2003. The impact of the certainty context on the process of choice. *Proceedings of the National Academy of Sciences, USA* 100(6): 3536–41.
Elliott, R., K. J. Friston and R. J. Dolan. 2000. Dissociable neural responses in human reward systems. *Journal of Neuroscience* 20(16): 6159–65.
Friedman, D., G. W. Harrison and J. Salmon. 1984. The Informational Efficiency of Experimental Asset Markets. *Journal of Political Economy* 92: 349–408.
Friston, K. J., J. Ashburner, C. D. Frith, J. B. Poline, J. D. Heather and R. S. J. Frackowiak. 1995. Spatial registration and normalization of images. *Human Brain Mapping* 2: 165–89.
Gallagher, H. L. and C. D. Frith. 2003. Functional imaging of 'Theory of Mind'. *Trends in Cognitive Sciences* 7: 77–83.
Garcia, C. B. and W. I. Zangwill. 1981. *Pathways to solutions, fixed points, and equilibria*. Englewood Cliffs, NJ: Prentice-Hall.
Gigerenzer, G. 1996. On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review* 103: 592–6.
Gilhooly, K. J. and W. A. Falconer. 1974. Concrete and abstract terms and relations in testing a rule. *Quarterly Journal of Experimental Psychology* 26: 355–9.
Glimcher, P. W. 2003. *Decisions, uncertainty, and the brain: The science of neuroeconomics*. Cambridge, MA: MIT Press.
Goeree, J. K., C. A. Holt and T. R. Palfrey. 2005. Regular quantal response equilibrium. *Experimental Economics* 8: 347–67.
Grether, D. M., C. R. Plott, D. B. Rowe, M. Sereno and J. M. Allman. 2007. Mental processes and strategic equilibration: An fMRI study of selling strategies in second price auctions. *Experimental Economics* 10: 105–22.
Gul, F. and W. Pesendorfer. 2006. Random expected utility. *Econometrica* 74: 121–46.
Gul, F. and W. Pesendorfer. 2008. The case for mindless economics. In *Foundations of positive and normative economics*, ed. A. Caplin and A. Schotter. New York: Oxford University Press.
Haile, P. A., A. Hortaçsu and G. Kosenok. 2008. On the empirical content of quantal response equilibria. *American Economic Review* 98: 180–200.

Harbaugh, W. T., U. Mayr and D. R. Burghart. 2007. Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316: 1622–5.

Harrison, G. W. 1989. Theory and misbehavior of first-price auctions. *American Economic Review* 79: 749–62.

Harrison, G. W. 1992. Theory and misbehavior of first-price auctions: Reply. *American Economic Review* 82: 1426–43.

Harrison, G. W. and J. A. List. 2004. Field experiments. *Journal of Economic Literature* 42: 1013–59.

Harrison, G. W. and E. E. Rutström. 2005. Expected utility theory *and* prospect theory: one wedding and a decent funeral. Working Paper 05-18. Department of Economics, College of Business Administration, University of Central Florida; *Experimental Economics*, forthcoming.

Harrison, G. W. and E. E. Rutström. 2008. Risk aversion in the laboratory. In *Risk aversion in experiments*, ed. J. C. Cox and G. W. Harrison. Bingley, UK: Emerald.

Harsanyi, J. C. 1975. The tracing procedure: A Bayesian approach to defining a solution for *n*-person non-cooperative games. *International Journal of Game Theory* 4: 61–94.

Hayes, J. R. and H. A. Simon. 1974. Understanding written problem instructions. In *Knowledge and cognition*, ed. L. W. Gregg. Hillsdale, NJ: Lawrence Erlbaum.

Hey, J. D. 2005. Why we should not be silent about noise. *Experimental Economics* 8: 325–45.

Hey, J. D. and C. Orme. 1994. Investigating generalizations of expected utility theory using experimental data. *Econometrica* 62: 1291–326.

Hirshleifer, J. and J. G. Riley. 1992. *The analytics of uncertainty and information*. New York: Cambridge University Press.

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

Johnson, E. J., C. F. Camerer, S. Sankar and T. T. Tymon. 2002. Detecting failures of backward induction: monitoring information search in sequential bargaining. *Journal of Economic Theory* 104: 16–47.

Johnson-Laird, P. N., P. Legrenzi and M. Sonino Legrenzi. 1972. Reasoning and sense of reality. *British Journal of Psychology* 63: 394–400.

Jones, L. S. 2007. The ethics of transcranial magnetic stimulation. *Science* 315: 1663.

Judd, K. L. 1998. *Numerical methods in economics*. Cambridge, MA: MIT Press.

Kahneman, D. and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263–91.

Kahneman, D. and A. Tversky. 1996. On the reality of cognitive illusions. *Psychological Review* 103: 582–91.

Keller, L., ed. 1999. *Levels of selection in evolution*. Princeton, NJ: Princeton University Press.

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

Knoch, D., A. Pascual-Leone and E. Fehr. 2007. The ethics of transcranial magnetic stimulation: Response. *Science* 315: 1663–4.

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.

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

Leland, W. J. 1994. Generalized similarity judgements: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty* 9: 151–72.

Linden, D. J. 2007. *The accidental mind*. Cambridge, MA: Harvard University Press.

Lohrenz, T., K. McCabe, C. F. Camerer and P. R. Montague. 2007. Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences, USA* 104: 9494–8.

Loomes, G. 2005. Modelling the stochastic component of behavior in experiments: some issues for the interpretation of data. *Experimental Economics* 8: 301–23.
Loomes, G. and R. Sugden. 1998. Testing different stochastic specifications of risky choice. *Economica* 65: 581–98.
Lopes, L. L. 1995. Algebra and process in the modeling of risky choice. In *Decision Making from a Cognitive Perspective*, ed. J. R. Busemeyer, R. Hastie and D. L. Medin. San Diego: Academic Press.
Luce, R. D. 1986. *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
Marr, D. 1982. *Vision: A computational investigation into the human representation and processing of visual information*. New York: W. H. Freeman & Company.
Marr, D. and T. Poggio. 1976. Cooperative computation of stereo disparity. *Science* 194: 283–7.
McCabe, K., D. Houser, L. Ryan, V. L. Smith and T. Trouard. 2001. A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences, USA* 98: 11832–5.
McClure, S. M., D. I. Laibson, G. Loewenstein and J. D. Cohen. 2004. Separate neural systems value immediate and delayed monetary rewards. *Science* 306: 503–7.
McClure, S. M., K. M. Ericson, D. I. Laibson, G. Loewenstein and J. D. Cohen. 2007. Time discounting for primary rewards. *Journal of Neuroscience* 27: 5796–804.
McDaniel, T. M. and E. E. Rutström. 2001. Decision making costs and problem solving performance. *Experimental Economics* 4: 145–61.
Mehta, J., C. Starmer and R. Sugden. 1994. The nature of salience: An experimental investigation of pure coordination games. *American Economic Review* 84: 658–673.
Merlo, A. and A. Schotter. 1992. Experimentation and learning in laboratory experiments: Harrison's criticism revisited. *American Economic Review* 82: 1413–25.
Montague, P. R. and G. S. Berns. 2002. Neural economics and the biological substrates of valuation. *Neuron* 36: 265–85.
Paris, Q. and M. R. Caputo. 2002. Comparative statics of money-goods specifications of the utility function. *Journal of Economics (Zeitschrift für Nationalökonomie)* 77: 53–71.
Pratt, J. W. 1964. Risk aversion in the small and in the large. *Econometrica* 32: 122–36.
Rabe-Hesketh, S., E. T. Bullmore and M. J. Brammer. 1997. The analysis of functional magnetic resonance images. *Statistical Methods in Medical Research* 6: 215–37.
Rabin, M. 1998. Psychology and economics. *Journal of Economic Literature* 36: 11–46.
Rilling, J. K., A. G. Sanfey, J. A. Aronson, L. E. Nystrom and J. D. Cohen. 2004. The neural correlates of theory of mind within interpersonal interactions. *NeuroImage* 22: 1694–703.
Ross, D. 2005. *Economic Theory and Cognitive Science: Microexplanation*. Cambridge, MA: MIT Press.
Rubinstein, A. 1988. Similarity and decision-making under risk (is there a utility theory resolution to the Allais paradox?) *Journal of Economic Theory* 46: 145–53.
Rubinstein, A. 2006. Discussion of 'Behavioral Economics'. In *Advances in Economics and Econometric Theory*, ed. R. Blundell, W. K. Newey and T. Persson, vol II, 246–54. New York: Cambridge University Press.
Rubinstein, A. 2007. Instinctive and cognitive reasoning: a study of response times. *Economic Journal* 117: 1243–59.
Samuelson, L. 2005. Foundations of human sociality: a review essay. *Journal of Economic Literature* 43: 488–97.
Samuelson, P. 1938. A note on the pure theory of consumer's behaviour. *Economica* 5: 61–71.
Sanfey, A. G., J. K. Rilling, J. A. Aronson, L. E. Nystrom and J. D. Cohen. 2003. The neural basis of economic decision-making in the ultimatum game. *Science* 300: 1755–8.
Simon, H. A. 1996. *The sciences of the artificial*, 3rd edn. Cambridge, MA: MIT Press.
Smith, V. L. 1991. Rational choice: the contrast between economics and psychology. *Journal of Political Economy* 99: 877–97.

Smith, V. L. 2003. Constructivist and ecological rationality in economics. *American Economic Review* 93: 465–508.

Smith, V. L. 2007. *Rationality in economics: Constructivist and ecological forms*. New York: Cambridge University Press.

Sorenson, R. A. 1992. *Thought experiments*. New York: Oxford University Press.

Starmer, C. 2000. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* 38: 332–82.

Stigler, G. J. and G. S. Becker. 1977. De Gustibus Non Est Disputandum. *American Economic Review* 67: 76–90.

Sugden, R. 2003. The logic of team reasoning. *Philosophical Explorations* 6: 165–81.

Sunder, S. 2006. Economic theory: structural abstraction or behavioral reduction? *History of Political Economy* 38: 322–42.

Tisserand, D. J., J. C. Pruessner, E. J. Sanz Arigita, M. P. J. van Boxtel, A. C. Evans, J. Jolles and H. B. M. Uylings. 2002. Regional frontal cortical volumes decrease differentially in aging: an MRI study to compare volumetric approaches and voxel-based morphometry. *NeuroImage* 17: 657–9.

Tversky, A. 1969. Intransitivity of preferences. *Psychological Review* 76: 31–48.

Tversky, A. 1977. Features of similarity. *Psychological Review* 84: 327–52.

Varian, H. R. 1985. Non-parametric analysis of optimizing behavior with measurement error. *Journal of Econometrics* 30: 445–58.

Varian, H. R. 1988. Revealed preference with a subset of goods. *Journal of Economic Theory* 46: 179–85.

von Winterfeldt, D. and W. Edwards. 1982. Costs and payoffs in perceptual research. *Psychological Bulletin* 19: 609–22.

von Winterfeldt, D. and W. Edwards. 1986. *Decision analysis and behavioral research*. New York: Cambridge University Press.

Wason, P. C. and P. N. Johnson-Laird. 1972. *Psychology of reasoning: structure and content*. Cambridge, MA: Harvard University Press.

Wason, P. C. and D. Shapiro. 1971. Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology* 23: 63–71.

White, N. M. and S. Gaskin. 2006. Dorsal hippocampus function in learning and expressing a spatial discrimination. *Learning and Memory* 13: 119–22.

Wilcox, N. T. 1993. Lottery choice: incentives, complexity and decision time. *Economic Journal* 103: 1397–417.

Wilcox, N. T. 2006. Theories of learning in games and heterogeneity bias. *Econometrica* 74: 1271–92.

Wilcox, N. T. 2008a. Stochastic models for binary discrete choice under risk: a critical primer and econometric comparison. In *Risk Aversion in Experiments*, ed. J. Cox and G. W. Harrison. Bingley, UK: Emerald.

Wilcox, N. T. 2008b. Stochastically more risk averse: a contextual theory of stochastic discrete choice under risk. *Journal of Econometrics* 142: forthcoming.

Wong, D. K., L. Grosenick, E. T. Uy, M. P. Guimaraes, C. G. Carvalhaes, P. Desain and P. Suppes. 2008. Quantifying inter-subject agreement in brain-imaging analyses. *NeuroImage* 39: 1051–63.

Zak, P. J., R. Kurzban and W. T. Matzner. 2005. Oxytocin is associated with human trustworthiness. *Hormones and Behavior* 48: 522–7.

Zak, P. J., A. A. Stanton and S. Ahmadi. 2007. Oxytocin increases generosity in humans. *PloS ONE*, 11: e1128, 1–5.