TWO STYLES OF NEUROECONOMICS

Don Ross

Economics and Philosophy / Volume 24 / Special Issue 03 / November 2008, pp 473 - 483
DOI: 10.1017/S0266267108002095, Published online: 05 November 2008

Economics and Philosophy, **24** (2008) 473–483
doi:10.1017/S0266267108002095

TWO STYLES OF NEUROECONOMICS

Don Ross
University of Cape Town and University of Alabama at Birmingham

I distinguish between two styles of research that are both called "neuroeconomics". *Neurocellular economics* (NE) uses the modelling techniques and mathematics of economics – constrained maximization and equilibrium analysis – to model relatively encapsulated functional parts of brains. This approach rests upon the fact that brains are, like markets, massively distributed information-processing networks over which executive systems can exert only limited and imperfect governance. Harrison's (2008) deepest criticisms of neuroeconomics do not apply to NE. However, the more famous style of neuroeconomics is *behavioural economics in the scanner*. This is often motivated by complaints about conventional economics frequently heard from behavioural economists. It attempts to use neuroimaging data to justify arguments for replacing standard aspects of microeconomic theory by facts and conjectures about human psychology. Harrison's grounds for unease about neuroeconomics apply to most BES, or at least to its explicit methodology. This methodology is naively reductionist and illegitimately assumes that economics should not do what all successful science does, namely, model abstract aspects of its target phenomena instead of would-be complete and fully ecologically situated facsimiles of them.

1. INTRODUCTION

Harrison's multi-pronged critique of neuroeconomics in the present symposium is a tour de force. Its best aspect, aside from its close attention to modelling and experimental design detail, is its disinterest in promoting the hegemony of any one ongoing research programme over another. Harrison even-handedly administers lumps to two groups: the would-be paradigm shifters who call for rebel attack on the Death Star of Imperial Orthodox Economics (e.g. Camerer, Loewenstein and Prelec 2005), and the "Hardnose Theorists" (Gul and Pesendorfer 2008) who really *do* defend the fortress after the fashion of Darth Vader. Harrison dissents from the Hardnosed view that the domain of economics is strictly limited a priori to

human rational choice, which would exclude from economists' terrain both the causal realm of brain events and the consumption behaviour of non-human animals. His questions are instead nicely pragmatic: Are the opportunity costs (in both intellectual capital and money) of most actual neuroeconomic experiments worth paying now?; and Can we see a path toward improving the cost–benefit ratio in the future? Harrison answers "no" to the first question, but clearly sees his main contribution as a general suggestion with respect to the second question. In this comment, I will argue that his constructive aim has been implicit from the outset in *one* main current within neuroeconomics. However, with respect to the other main current – the current now generating most of the overblown claims against which Harrison reacts – economists who wish to avoid cluttering the disciplinary agenda should indeed be very cautious.

2. NEUROCELLULAR ECONOMICS

There are several kinds of activities going on under the rubric of "neuroeconomics". The most informative broad cut, I suggest, is between what I will call "behavioural economics in the scanner" (BES) and "neurocellular economics" (NE). BES mainly consists of repeating protocols from behavioural economics experiments – ultimatum games, Prisoner's Dilemmas, intertemporal choice paradigms to elicit discounting patterns, etc. – while subjects' brains are indirectly observed under neuroimaging.[1] NE is the programme of using the mathematics of economic equilibrium analysis to write down models of brain cell activity for the sake of refining and comparatively testing hypotheses about neural learning that originate from computational neuroscience. Most of the recent and recurrent excitement in popular media over neuroeconomics has focused on experiments in the BES vein, and this attention is playing a major role in facilitating economists' access to scanning equipment at one institution after another.[2] NE, on the other hand, retains a clear geographical and intellectual centre of gravity at New York University, with a major descendent outpost at Duke, and a cousin at Baylor College of Medicine, and lives beneath the radar of most economists. This is despite the fact that the first high-profile publication in neuroeconomics, Glimcher (2003), is the basic methodological statement of NE.

The motivations of NE are easier to discern, easier to straightforwardly describe and, in my view, easier to defend in an unconflicted spirit

[1] One of Harrison's many valuable services in his paper is emphasizing just how highly indirect the observations in question are. Neuroimaging textbooks, such as Huettel *et al.* (2004) are clear about this. Surveys intended mainly for economists should be too, but usually aren't.

[2] Disclaimer: this comment squarely applies to the research teams of which I am a funded member at both universities that employ me.

than is the case for BES. The mammalian brain controls behaviour by learning about contingent relationships between environmental predictors of reward and classes of actions. For trying to understand neural learning, the method of constructing computational simulations dominates all others (Sutton and Barto 1998). However, this approach is highly abstract from the biological perspective: computational learning theory is intended to apply to many sorts of systems besides brains. For example, it is also intended to apply to markets learning expected asset values from observed movements in prices. One important property that brains and markets have in common as computers is that both are massively parallel processors.[3] This doesn't necessarily matter to the descriptions "in extension" (as philosophers say) of the functions they compute.[4] But it matters a great deal to models of their learning processes. We know of only one kind of technology for this job: models in which functional forms and values are estimated for (local or global) optimization of some or other parameter or weighted set of parameters. Learning is (locally or globally) accomplished when such processes reach equilibrium. Thus, as McCabe (2008, this issue) says, a bit vaguely, "equilibrium and computation are organized by the general principle of optimization".

Abstract theory in this area has been thriving for decades. Economists interested in the sculpting of expectations among agents who seek optimal mixes of savings and investments in response to price changes and macroeconomic policy signals have assembled a massive body of achievement, exhaustively axiomatized (as of his point of writing) by Stigum (1990). But no one ever imagined that brains are *general* implementations of this abstract theory. Simulations of specifically *neural* learning add biologically inspired constraints, and apply these to specific sub-regions of the brain on the basis of evidence that different sub-regions are responsible for computing different functions in partial encapsulation from one another. A particularly well-articulated model of the midbrain/striatal dopamine circuit (Montague *et al.* 1996; Schultz *et al.* 1997; Montague and Berns 2002) depicts a system that integrates attentional cueing, value estimation and motor response preparation: a veritable engine room for consumption. Daw (2003) simulates this, and into the bargain merges it with the behavioural observations theoretically consolidated by Gallistel (1990) and Gallistel and Gibbon (2000) suggesting that "raw" animals (i.e. animals not assisted by prostheses such as institutionalized price records) approximate behavioural optimization by

[3] Brains in addition have forms of executive guidance (Christensen 2007). Watching the US Federal Reserve hyperactively trying to keep credit markets optimistic at the time of the present writing (March 2008), one is reminded that some markets may as well.

[4] This refers to a function characterized simply as a class of input-output mappings, which does not distinguish among procedural variants.

choosing environments that maximize reward rates over learned temporal intervals.

Simulation is just what the doctor orders for answering "how possibly?" questions, and, more importantly, for elaborating models by discovering how their elements dynamically interact under varying parameterizations. Still, if our models are supposed to be proto-neural applications, they need to be connected to neural data. This is where NE, as pioneered by Glimcher and his students, enters the picture. The basic method is to (i) incentivize an animal to learn the predictive value of a reward cue, (ii) derive predicted changes in neural firing rates that should occur *somewhere* in a conjectured brain region (of larger or smaller scope depending on prior knowledge) if the modelled learning is to be implemented along a path terminating in equilibrium, then (iii) look for the changes in question while monkeys or people are scanned under fMRI (Platt and Glimcher 1999). Exemplifying the fruits of this approach, Glimcher (2003) interprets changes in firing rates in parietal cortex while monkeys play inspection games as tracking Nash equilibrium, and Glimcher *et al.* (2007) report patterns of striatal firing in a human intertemporal choice task that they take as predicting revealed preferences over rewards at varying delays.

The second protocol above is critically discussed (including by the authors) in the context of a miniature tradition started by the famous McClure *et al.* (2004) paper on putative dual discounting systems in the brain that is criticized by Harrison. Notwithstanding the fact that Glimcher *et al.* take their data to support McClure *et al.*'s null hypothesis, the underlying purpose of the two experiments is importantly different. McClure *et al.* present neural evidence as a basis for a specific theory of human divergence from standard economic rationality with respect to a basic aspect of consumption behaviour, discounting. By contrast, the Glimcher *et al.* experiment is motivated primarily in service of the development of a computational model of the reward system. Caplin and Dean (2007), Glimcher's colleagues in NYU's Economics Department, are directly on this job, setting out to axiomatize the dopamine circuit as an economic agent in light of the results from across campus (and elsewhere, of course). In so doing, they see themselves as explicitly responding to Gul and Pesendorfer's challenge – though, if they are successful, Hardnose critics might argue that they better demonstrate the relevance of economics to neuroscience than the other way around.

Three points about NE are important in the context of Harrison's critique.

First, the only main concern of his that seems potentially damaging to it, that its typically small sample sizes interact problematically with frequently naïve econometrics used for testing interpretations of data, is a complaint that (i) is the daily bread and butter of criticism across

the length and breadth of economics, and (ii) admits of straightforward solution, given big enough research budgets.

Second, we should not expect the NE approach to be specifically vulnerable (i.e. more vulnerable than other sciences that rely on elaborate statistical analysis for obtaining observations) to deep philosophical criticisms of the kind Harrison raises for BES. The kinds of models NE researchers seek are the stock in trade of all computational neuroscience, supplemented by experimental economists' acquired wisdom on the difficulties involved in experimentally isolating subtly different optimization hypotheses.

Third, NE is a zone in which we should be particularly sceptical in response to disciplinary moat digging by Hardnose theorists. I see little profit in trying to find the border between general theories of learning in markets and theories of learning in wide classes of systems in cognitive science. Lionel Robbins (1935) taught us long ago that the basic subject matter of economics is minimization of opportunity costs by creatures to which we have a sound basis for imputing objective functions.[5] And now that we have giant computers we needn't usually rest content with comparative statics. It is thus no surprise that each passing year sees more of the graduate curriculum in a typical economics department taken up by dynamic optimization techniques. One doesn't have to climb very far up the ladder of abstraction here before one notices that computational learning theorists and economists are preoccupied with the same objects of study: Markov and ARIMA processes, stochastic difference learning, non-stationarity, etc. – and, in Glimcher's case, games. Of course, economists are interested in subjects other than learning, and some learning theorists study problems – e.g. an animal's remembering where its nest is – that don't typically distract economists. But the NE theorist essentially borrows economic theory to write down models of neural reward learning. Economic theory is the right theory to borrow because, dynamically, reward valuation just *is* learning and the brain must somehow pull it off.

In this light, I will explain why I refer to the NYU style of neuroeconomics as "neurocellular economics".

McCabe (2008, this issue) comments that "conceptually, a neural system is very similar to a microeconomic system, but the process of designing an experiment to study neural function is quite different since a neural system does not define an agent to which a subject can be mapped." This seems to presuppose that an economist who specifies a microeconomic system must be assumed to be clear about agency. In one crucial sense this is true: in economic models, agents just *are*

[5] He thought that only people are such creatures. However, his grounds for this don't pass muster in cognitive science (Ross 2005; 87–99).

utility and production functions, and at least one such function is at least implicitly specified in any economic model. However, I suspect that this isn't what McCabe has in mind; I read him instead as suggesting, with many anti-establishment proponents of BES, that the identities of the agents economists model are assumed to be fixed independently of their modelling decisions, by common sense or by other sciences such as psychology, biology or anthropology. I think this assumption is false.

One way to understand the incessant complaining about Death Star economics that we hear from philosophers, champions of "humanism" among economists such as Sen (1977, 1987), and "rebel" behavioural economists, is that in associating agency with utility functions we ignore most of the interesting characteristics of agents. What is really meant by this objection is that we ignore most properties of whole *people*; it is taken for granted that the canonical economic agent *should* be a whole person. This is *essential* to the non-disingenuous failure of many behavioural economists to notice that their anti-establishment rhetoric is, as Harrison says "lousy, sucker-punch economics". But the assumption has no rigorous basis. (See Ross 2005 for a detailed review, and Ross forthcoming for an update.) How many working economists *ever* try to construct the utility function of a whole person, as opposed to (e.g.) the utility function of a representative household over a limited set of consumption goods this year, or the utility function of a representative wage-earner with respect to her distribution of consumed income versus saved wealth over so many years, etc.? That economists abstract away from most psychological properties of people is a stick with which the discipline is often beaten, but this is grotesquely unreasonable. *All* scientific modelling represents selected aspects of phenomena; that is precisely the point of modelling. Someone could coherently argue – against a vast weight of evidence to the contrary – that economists never show us anything useful when they model situations involving humans. But the entirely appropriate response to being told we shouldn't model *selected* aspects of human behaviour in the first place is just a hard stare.

Thus, economic agents are constructed by economists in the course of modelling. A good part of the difference in skill between a better and a worse economist is having a feel, learned on the job, for what sort of agents it's a good idea to construct as part of the task of building a model of a situation that will prove to be robust and insightful and survive econometric testing. I doubt that there is any deep philosophical story waiting to be unearthed about this; it is a kind of abstract perception.

Once we recognize that economics observes "metaphysical minimalism" with respect to agency – that is, that an agent arises anywhere a utility or production function pops up in a good model – then light is shed on why NE is on clearer methodological ground than BES. Once a modeller has gotten as far as conceiving of a neural system as learning by

optimizing over a reward space, there is automatically an implicit agent in the picture, and its identity is clear enough for practical purposes[6] – it is *that* neuron, or *that* neurotransmitter system, or *that* quasi-modular circuit.[7] Achievement of equilibrium – completion of learning – by the neuron or circuit may often (though often not) also maximize imputed utility of the larger organism, and we had better ultimately have *some* story to tell about this relationship in any given case; but this is an entirely contingent matter from one instance to the next. Thus NE is the *economics* of the behaviour of *neurocellular agents*.

Harrison comments that "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?" Precisely so. This point can now be applied to explain why BES is generally on much shakier methodological ground than NE.

3. BEHAVIOURAL ECONOMICS IN THE SCANNER

In my view the two most important sentences in Harrison's paper (the second of which is a quote from Sunder) are the following:

> The algorithmic approach offers another model of the 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.
> There is no internal contradiction in suboptimal behavior of individuals yielding aggregate-level outcomes derivable from assuming individual optimization. Individual behavior and aggregate outcomes are related but distinct phenomena [Sunder].

We are all very used to the idea that rational, informed optimizers can, without committing any mistakes, interact in ways that earn them Pareto-inferior payoffs. Economists are equally familiar with scenarios in which quite stupid people, who can't grasp that trade isn't a zero-sum game, or that frequency differences between two variables are only informative given base rates, or etc., operate with easy success in intricately complicated networks of market relations. This was the basis of Milton Friedman's (1953) classic injunction to economists to forget about the realism of their assumptions, and model people in markets *as if* they were computers of rational optima. Unfortunately, Friedman had no plausible philosophical rationale to accompany this advice. He cut himself off from the possibility of such a story by assuming (along with nearly every

[6] It won't be clear enough for philosophical purposes – but nothing ever is.

[7] Harold Kincaid points out to me that identity conditions on what neuroscientists call "circuits" are anything but generally clear. However, this is not the source of the "undefined agency" to which McCabe aims to refer.

other sensible thinker in the 1950s)[8] reductionism: "high-level" entities like markets must decompose by steps into physical atoms.

But this 1950s orthodoxy turned out not to be true, either metaphysically or, more relevantly in the present instance, socially. Where economics, psychology and neuroscience are concerned, reductionism fails on at least two levels. First, people's distributed solutions to their local coordination problems tend to converge to culturally shared icons – which are often regimented into institutions – that store *collective* learning about what it's sensible to do in different circumstances. This collective learning needn't be explicitly encoded by any of the individuals, who need only develop the habit of following the cultural rule. Individuals can get by with less rationality the more they can rely on what Clark (1997) calls "scaffolding" in their cultural and institutional environments (just as they can hunt successfully with degraded olfaction and hearing if they have dogs). Second, and by exactly parallel logic, whole brains can store problem solutions in a distributed way, visible only in molar behaviour, that no molecular part of the brain in question encodes (Clark 1989; Dennett 1991).

As emphasized by Satz and Ferejohn (1994), the general, non-mysterious failure of reductionism in social science has immediate implications for economics. For example, macroeconomic regularities can explain some microeconomic ones, rather than the other way around. The microfoundations tradition may be a useful way of introducing the technical rigor of microeconomics into macroeconomics, but one should not therefore imagine that real participants in macroeconomies approximate representative agents. From the fact that people reverse preferences in laboratory experiments one should not expect markets to react chaotically to changes in supply and demand.

Anti-establishment manifestos inspired by reports from BES, such as elicit Harrison's justifiably caustic reaction, are typically driven by reductionist assumptions. If people's dopamine systems react to money just as they do to ice cream, then there is a rush to assume that people will behaviourally respond to changes in the money supply just as they respond to changes in the ice cream supply, and all economic models that don't incorporate diminishing marginal utility of money are to be rejected. Knutson *et al.* (2007) comment with straight faces that their observations of dopamine response in subjects choosing purchases in their lab indicate that people don't respond to opportunity cost, so economic theory requires fundamental revision (to put it mildly). Harrison rightly corrects the attribution to "economists" by Camerer *et al.* (2005) of the assumption that people maximize hedonic pleasure, but in fact these authors stand in need of a double correction: even if brains *did* light up like jackpots when their owners had orgasms while staying quiet as mice when their owners'

[8] But not the greatest of them, von Neumann.

asset portfolios appreciated, this wouldn't necessarily imply anything at all about how economists should write down the subjects' utility functions.

In often arguing in this way, BES-style neuroeconomists simply repeat the logical mistake of rebel behavioural economists. However, for the reasons I indicated above, they compound it by assuming reduction at two levels instead of just one: first, from the institutionally embedded person to the de-contextualized lone mind in the lab (the original behaviourist slip), and then from the lone mind in the lab to the de-contextualized lone neurotransmitter system in the head. The reductionistic assumptions lead to failure to distinguish enough economic agents. When we apply NE, and explicitly or implicitly construct a utility function for a neural processing system, we should not assume that that utility function also characterizes the molar agent, i.e. the whole subject over the course of an experiment or other consumption episode.

Note that, despite Harrison's recognition near the end of his paper that problems over agency location are close to the root of the trouble, the main complaint running through his opening gallery of instances of dodgy reasoning in neuroeconomics is not, per se, that the guilty parties try to construct subjects' utility functions directly from interpretations of neural data. Rather, his gripe is that they use the data to refute straw men derived from inaccurate and uncharitable representations of Dark Star economics. The point about reductionism and agency, however, is key to *explaining* this tendency to misrepresentation. If utility *did* have to map directly on to something one could observe in people's heads, then someone might naturally have expected the something in question to be a pleasure signalling reflex until neuroscientists came along to distinguish "wanting" from "liking" at the neural level. Similarly, *because* it is assumed that economists must have intended state-independent time preferences to be based on some to-be-observed neural register, it is taken to spell trouble for them when glimpses into the brain reveal no such gadget.

In his essay in the present symposium, McCabe does not call for any sorties against the Death Star. Thus neither Harrison's brickbats, nor my complaints above, are called down upon him. On the other hand, he can frequently be read in a way that invites the hasty inferences typical of BES-style neuroeconomics.

For example, at one point McCabe says "Humans invent institutions, such as money, to facilitate our goal-directed behaviours and these institutions in turn get instantiated in the brain." The meaning of this is ambiguous. It can be interpreted as meaning only that since people learn to prefer more money to less, and to trade things for money in consistent ways, their brains must be able to process information about monetary values, and to sort different kinds of perceptual events into equivalence classes in units of money. On this reading, the comment is surely true. Furthermore, it might indeed be relevant to economists to

learn some of the neuroscience of this processing. (Suppose, for example, that brains systematically perceive "$14.99" as more similar to "$14.00" than to "$15.00". Then there might be discontinuities between people's comparative evaluations of pairs of small purchases and their comparative evaluations of pairs of large ones, and we should be cautious about using data on the former to reveal preferences over the latter.)

However, McCabe can also be read as suggesting that across individual brains we should find, as it were, a fractal reproduction of the logic of the institution of money. This would get us as far as the bi-directional causation to which Harrison refers: individuals would coordinate their behaviour to create money (micro-to-macro causation), then each individual would form a private representation of the new institution (macro-to-micro causation), appropriate (micro) behavioural patterns could be generated in each individual, and the result is meaningful monetary prices at the aggregate scale. However, the real causal network is more complicated than that. Brains need not "instantiate" many properties of the institution of money that nevertheless *are* properties of that institution, and which influence individual patterns of behaviour. For example, none need recognize that an increase in current borrowing lowers the value of current units; they need merely be cued to copy the behaviour of whoever they see setting the highest price and still getting a sale. Then if a lot of them borrow but the supply of goods and services stays constant they will get inflation. The difference between narrow and broad money can be causally relevant to behaviour even if it is "instantiated" in no brains but those of economic analysts.

When I say that McCabe "can be" read this way, I don't mean to uncharitably suggest that he imagines we'll ultimately find in each consumer's brain a complete guide to commercial and social institutions. My point, rather, is that it is easy to *lazily* conceptualize the representation of institutions and economic patterns by brains in a way that fails to expose fallacies of reductionism and reification of agents at different scales. Much of the neuroeconomics literature that starts directly from the behavioural economics literature and then adds neuroimaging is a veritable Mississippi River of such reasoning. Two elegant edifices of formal construction, economics and computational learning theory, deserve to *both* be treated with more respect when they're brought together. Fortunately, in the literature of NE they generally are.

REFERENCES

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

Caplin, A. and M. Dean. 2007. The neuroeconomic theory of learning. *American Economic Review* 97: 148–52.

Christensen, W. 2007. The evolutionary origins of volition. In *Distributed cognition and the will*, ed. D. Ross, D. Spurrett, H. Kincaid and G. L. Stephens, 255–87. Cambridge, MA: MIT Press.
Clark, A. 1989. *Microcognition.* Cambridge, MA: MIT Press.
Clark, A. 1997. *Being There.* Cambridge, MA: MIT Press.
Daw, N. 2003. *Reinforcement learning models of the dopamine system and their behavioral implications*. Doctoral dissertation, Carnegie Mellon University, Pittsburgh.
Dennett, D. 1991. *Consciousness explained*. Boston: Little Brown.
Friedman, M. 1953. *Essays in positive economics.* Chicago: University of Chicago Press.
Gallistel, C. 1990. *The organization of learning.* Cambridge, MA: MIT Press.
Gallistel, C. and J. Gibbon. 2000. Time, rate and conditioning. *Psychological Review* 107: 289–344.
Glimcher, P. 2003. *Decisions, uncertainty and the brain*. Cambridge, MA: MIT Press.
Glimcher, P., J. Kable and K. Louie. 2007. Neuroeconomic studies of impulsivity: now or just as soon as possible? *American Economic Review* 97: 142–7.
Gul, F. and W. Pesendorfer. 2008. The case for mindless economics. In *The foundations of positive and normative economics*, ed. A. Caplin and A. Schotter, 3–41. Oxford: Oxford University Press.
Harrison, G. W. 2008. Neureconomics: a critical reconsideration. *Economics and Philosophy* 24.
Huettel, S., A. Song and G. McCarthy. 2004. *Functional magnetic resonance imaging*. Sunderland, MA: Sinauer.
Knutson, B., S. Rick, G. E. Wimmer, D. Prelec and G. Loewenstein. 2007. Neural predictors of purchases. *Neuron* 53: 147–56.
McCabe, K. A. Neuroeconomics and the economic sciences. *Economics and Philosophy* 24.
McClure, S., D. Laibson, G. Loewenstein and J. Cohen. 2004. Separate neural systems value immediate and delayed monetary rewards. *Science* 306: 503–7.
Montague, R. and G. Berns. 2002. Neural economics and the biological substrates of valuation. *Neuron* 36: 265–84.
Montague, R., P. Dayan and T. Sejnowski. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience* 16: 1936–47.
Platt, M. and P. Glimcher. 1999. Neural correlates of decision variables in parietal cortex. *Nature* 400: 233–8.
Robbins, L. 1935. *An essay on the nature and significance of economic science*, 2nd Edn. London: Macmillan.
Ross, D. 2005. *Economic theory and cognitive science: Microexplanation.* Cambridge, MA: MIT Press.
Ross, D. forthcoming. The economic agent: not human, but important. In *Handbook of the Philosophy of Science*, Vol. 13. *Economics*, ed. U. Mäki. Amsterdam: Elsevier.
Satz, D. and J. Ferejohn. 1994. Rational choice and social theory. *Journal of Philosophy* 91: 71–87.
Schultz, W., P. Dayan and P. R. Montague. 1997. A neural substrate of prediction and reward. *Science* 275: 1593–9.
Sen, A. 1977. Rational fools. *Philosophy and Public Affairs* 6: 317–44.
Sen, A. 1987. *On ethics and economics.* Oxford: Blackwell.
Stigum, B. 1990. *Toward a formal science of economics*. Cambridge, MA: MIT Press.
Sutton, R. and A. Barto. 1998. *Reinforcement learning: An introduction.* Cambridge, MA: MIT Press.