Economics and Philosophy, **24** (2008) 459–471
doi:10.1017/S0266267108002083

FROM COGNITIVE SCIENCE TO COGNITIVE NEUROSCIENCE TO NEUROECONOMICS

Steven R. Quartz
California Institute of Technology

As an emerging discipline, neuroeconomics faces considerable methodological and practical challenges. In this paper, I suggest that these challenges can be understood by exploring the similarities and dissimilarities between the emergence of neuroeconomics and the emergence of cognitive and computational neuroscience two decades ago. From these parallels, I suggest the major challenge facing theory formation in the neural and behavioural sciences is that of being under-constrained by data, making a detailed understanding of physical implementation necessary for theory construction in neuroeconomics. Rather than following a top-down strategy, neuroeconomists should be pragmatic in the use of available data from animal models, information regarding neural pathways and projections, computational models of neural function, functional imaging and behavioural data. By providing convergent evidence across multiple levels of organization, neuroeconomics will have its most promising prospects of success.

Many neuroscientists incorporating economic theory and methods into their research would find the intuitions behind Stanley Jevons's attempt of 1879 to root economics in the materialist psychophysiology of his day surprisingly familiar (Jevons 1871, Maas 2005). Based on psychophysical attempts to discover quantitative laws of sensation, Jevons's unflinchingly materialist programme of a "mechanics of utility and self-interest" bear a striking resemblance to the same research traditions that led to both modern neuroscience and to the contemporary project of discovering the neural mechanisms underlying economic behaviour. It may not be too great an overstatement to note that in one direction an application of the Fechner–Weber law led to the notion of diminishing marginal utility while in the other direction it led to modern neuroscience in the union of visual psychophysics and neurophysiology (Kuffler 1953, Hubel and Wiesel 1959). Whatever intriguing parallels there may have been between the historical roots of neoclassical economics and modern neuroscience, however, these fundamentally diverged when economics took its "Paretian turn" (Bruni and Sugden 2007). Pareto's attempt to eliminate all psychological notions from the foundation of economics, replacing laws of sensation with ones regarding abstract choice or "logical action", later completed by Hicks, Samuelson and others in the theory of revealed preference, ultimately made economics a separate, autonomous and thus supposedly irreducible domain of inquiry.

As this volume illustrates, a century and a quarter after Jevons's *Theory of Political Economy*, a nascent neuroeconomics is seeking to re-ground economic behaviour in the current-best materialist understanding of human behaviour, cognitive neuroscience. Given the structure of neoclassical economics and its core methodological commitments, it is not surprising that this project has generated considerable debate, commentary and critique. While many of these stem from the question of how neuroeconomics stands in relation to neoclassical economics, others are the inevitable consequence of new interdisciplinary enterprises, as distinct disciplinary practices, methods, and standards come into contact and require integration.

The debates and controversies neuroeconomics has sparked strike me as reminiscent of those in an earlier episode in the neural and behavioural sciences, one that occurred approximately 20 years ago. In what follows, I want to explore the similarities – and dissimilarities – between this intellectual episode and the emergence of neuroeconomics, as a sort of historical case study, which I believe can shed light on the current state of neuroeconomics, its prospects and its challenges.

THE EMERGENCE OF COGNITIVE NEUROSCIENCE

The episode I have in mind is the emergence of cognitive and computational neuroscience from cognitive science, beginning in the early to mid-1980s. Before doing so, it is necessary to recap the broadest contours of cognitive science and its methodological commitments. The foundations of cognitive science were formed in the 1950s as cognitive psychologists (e.g. George Miller, Jerome Bruner), artificial intelligence and computer scientists (e.g. John McCarthy, Marvin Minsky) and linguists (e.g. Noam Chomsky) reacted against behaviourism by positing that the mind was a physical symbol processor. According to this view, the mind's operations could be characterized at an abstract level of description, the semantic and algorithmic, which in turn could map onto multiple physical implementations, later codified as a functionalist approach to mind (Putnam 1975). So construed, the semantic level of description, under which cognition was understood as the manipulation of symbols with propositional content according to the principle of rationality, was an autonomous level of description (Pylyshyn 1984). The reason for the autonomy of this level of description was in the claim that there were generalizations at the semantic level (in terms of propositional attitudes such as beliefs) for which there was no unitary physical-level description. As a consequence, cognitive science was argued to be irreducible to the physical sciences. In particular, since the semantic level of description could be implemented in different substrates, neural-level evidence seemed largely irrelevant to the explanation of mind.

While cognitive science was a productive research programme, it was not without limitations, some of which in retrospect were fundamental. For example, its research strategy of mapping cognitive tasks onto computational models proved to be severely under-constrained. For any set of behavioural data (e.g. a working memory task) there were many behaviourally equivalent computational models. The problem of model selection was compounded by the fact that the entire class of models being utilized incorporated unrealistic assumptions, such as unbounded resources, which in turn were justified on the basis of distinctions, such as performance/competence in linguistics, that struck many as problematic. By the mid-1980s, this functionalist cognitive science began to be challenged by a group of researchers in connectionism or parallel distributed processing (Rumelhart and McClelland 1986). These researchers challenged the autonomy of cognitive science from the physical sciences, and particularly cognitive science's central claim that the level of implementation was irrelevant to the explanation of cognition and behaviour. In particular, connectionism embraced a model of computation that had its roots in the perceptron linear classifiers of Rosenblatt (1958) and others. Whereas neural network research had been unpopular due to the computational limitations (Minsky and Papert 1969) explored for the case of Rosenblatt's single-layer perceptrons (and incorrectly conjectured would hold for more complex networks), the discovery, or re-discovery as the history is complex (Werbos 1994), of the back-propagation learning algorithm created tremendous excitement over the prospect of 'brain-like' computation.

Beginning around 1986, there was a dramatic realignment of research in the neural and behavioural sciences. The reasons for this are complex, but a few key points are worth highlighting here. First, neural computation provided important explanatory links between the high-level symbolic explanations of cognitive psychology and the low-level, mechanistic explanations of neuroscience. Thus, psychologists working in areas as diverse as visual psychophysics, memory, and learning were drawn to these modelling efforts as a way to constrain explanation. This led, ultimately, to cognitive neuroscience, which has in a remarkably short time become the standard approach to cognition (Gazzaniga 2004). In the other direction, neuroscience had amassed a tremendous amount of data regarding single neurons, from the dynamics of single ion channels and receptors using patch clamp techniques to the level of receptive field properties, but had few techniques or methods to integrate this information into theories of neural circuit function, large-scale theories of brain function, or even detailed models of single-neuron processing, such as dendritic integration. Computational modelling offered a way to integrate this information into such large-scale theories, and was thus rapidly integrated into neuroscience, in fact, creating a new sub-field of computational or theoretical neuroscience (Dayan and Abbott 2001).

THE CRITIQUE FROM WITHIN: NEOCLASSICAL ECONOMICS' REACTION TO NEUROECONOMICS

As an outsider looking into the debate within economics, many of the debates between what I will refer to loosely as mainstream economics and the proponents of neuroeconomics strike me as deeply reminiscent of those between cognitive science and connectionism. For example, the proponents of "mindless" economics (Gul and Pesendorfer 2008) charge neuroeconomists with making a category mistake by supposing that neural evidence could in principle be relevant to economics, as they "address different types of empirical evidence".

A similar argument was made by cognitive scientists against the relevance of neural evidence by claiming that neural and symbolic levels of description were autonomous, so any attempt to constrain one level of description by evidence from another amounted to a conceptual confusion (Pylyshyn 1984). As a graduate student during the 1980s working on the relationship between cognitive and neural development, I can recall psychologists looking quizzically at my suggestion that cognitive processes described as learning could participate in the construction of the neural circuits underlying that learning (Quartz and Sejnowski 1997). One pointed out that the suggestion was a non-starter, as Chomsky had stipulated an a priori "stationarity" principle making this impossible. Many of the claims made by proponents of mindless economics regarding the autonomy of economics have the same flavour of the Chomskyian whose commitments to model-theoretic principles are so severe that they rule out entire domains of potential empirical support or disconfirmation. Indeed, it is hard to otherwise understand how Gul and Pesendorfer (2008) could make the claim that "neuroscience evidence cannot refute economic models because the latter make no assumptions and draw no conclusions about the physiology of the brain" other than by invoking a strong autonomy thesis. While economic theory may make no *explicit* assumptions about the physiology of the brain, it is not the case that it makes no predictions that could either be confirmed or disconfirmed by neuroscience. For example, financial decision theory specifies the minimal parameters necessary for rational choice under uncertainty, expected reward and risk as variance of reward (Markowitz 1952). While financial decision theory makes no explicit predictions regarding neural correlates of these parameters, it does by implication make the testable prediction that neural activity will correlate separately with these two parameters given the appropriate task. A possible rejoinder is the claim that the theory does not make such a prediction since the function of theory is not to postulate terms that may or may not correspond to entities in the world, but rather to provide *as if* explanations whose sufficiency is determined by instrumental considerations, such as predictive capacity (Friedman 1953). While a consideration of realist vs. instrumentalist conceptions of economic methodology is beyond the scope of this paper (Lagueux 1994), it is worth noting that instrumentalist interpretations of the symbolic level of cognitive science (Dennett 1987) are generally regarded as failed programmes, which may lack internal coherence (Baker 1989). Furthermore, while an instrumentalist may deny that a correspondence (or lack thereof) between theoretical terms and real-world entities is the proper measure of a theory, the issue of such correspondence is nonetheless still an empirical one that may be tested.

In a recent study, we investigated this issue and found that brain activation in human striatum correlated with reward and risk and were differentiated both spatially and temporally and arose in the absence of learning, motivation or salience confounds (Preuschoff *et al.* 2006). To us, it was striking that the implicit predictions of financial decision theory were confirmed in neural activity. There was no a priori reason why this should be the case, and such a confirmation strengthens the theory beyond that provided by an *as if* instrumentalism.

The more general point I want to make is that the autonomy thesis of cognitive science proved to be a major barrier to progress in that field. It created a methodological isolationism within cognitive science that resulted in its failure to incorporate neuroscience into its approach, even though the case for such integration should have been clear. The result was not only that its theories remained highly under-constrained and it lacked a means to adjudicate among multiple equivalent theories, but it created an intellectual inflexibility that ultimately led to its obsolescence. If there is a cautionary tale for mindless economics contained in this historical episode it is that strong autonomy claims often lead to methodological isolationism and inflexibility, particularly in the face of growing methodological alternatives.

It may be worth noting that one reason for this methodological isolationism is often the perception that the integration of evidence from another field will be primarily disconfirming. Connectionists presented their models primarily in this disconfirming mode. In particular, connectionists argued that the syntactic transformations that cognitive science supposed were the core operations of a symbol system (a finite set of syntactic transformation rules applied to a finite set of symbols is essentially the definition of a grammar and language of thought, respectively) would be eliminated in a connectionist scheme (Rumelhart and McClelland 1986). While this claim may have been an important part of the revolutionary fervor of connectionism, in retrospect it was a relatively minor element of the programme with increasing links between connectionist models and semantic levels of representation (Rogers and McClelland 2004). Had cognitive science been more methodologically pragmatic, it likely could have absorbed these new developments rather than being displaced. Much neuroeconomics is likewise presented as primarily disconfirming to mainstream economics. Whether this will in fact be the case remains to be seen.

NEUROECONOMICS WITHIN NEUROSCIENCE

As the lead papers illustrate, a major driving force in the formation of neuroeconomics has been by behavioural economists looking to neuroscience to inform and constrain economic theory. In some respects, this accords with a similar trajectory of cognitive psychologists toward integrating neuroscience over the last few decades, who increasingly looked to neural constraints on behavioural data. However, another driving force that is relatively unexplored in the lead papers is in the other direction, namely from neuroscientists, who are eager to incorporate portions of economic theory into neuroscience. In many respects, the integration of economics into neuroscience is a natural extension of the incorporation of computational models into neuroscience in the late 1980s. In particular, a result of that integration was a major shift in how neuroscientists think about the brain and the organization of behaviour more generally. In cognitive science, the dominant view of the mind was as an abstract symbol processor, whose state transformations coincided with cognition as exemplified in such capacities as language production and problem-solving. The more basic view that minds/brains evolved as systems to navigate environments according to satisfying needs, that is, to find reward and avoid punishment, was marginalized in cognitive science. This was probably due to the fact that cognitive science grew out of a rejection of behaviourism, as exemplified in Chomsky's review of Skinner's *Verbal Behavior* (Chomsky 1959). To this day, discussions of reward learning and conditioning are still often met by cognitive psychologists suspiciously as attempts to reintroduce behaviourist notions that had, according to the view, been dismissed long ago.

Within neuroscience, however, by the early 1990s computational accounts of reinforcement learning provided the theoretical insights necessary to develop a novel approach to the function of a key neural system that until then had been intensively studied but poorly understood. Specifically, the midbrain dopamine system, a collection of small clusters of cells that project widely to the cortex, had long been implicated in reward, motivated behaviour and addiction, but the specific functional role of this system remained a mystery (Koob and Swerdlow 1988). In part, this was due to the fact that, unlike many other neural systems, the projections of this system were diffuse and non-specific, making it problematic to know how such a system could be involved in anything more than altering a relatively global feature of neural processing, such as gain. The temporal difference model of reinforcement learning required such a signal to broadcast a valuation signal for reward learning (Sutton and Barto 1998). As this model was applied to the mammalian midbrain dopamine system and to homologous systems in insects, it demonstrated that these reinforcement learning algorithms were computationally more powerful than behaviourist ones, bringing about an important shift in thinking about the centrality of reward processing to neural systems (Montague *et al.* 1995, 1996). This was central to a rapid growth of interest among neuroscientists more generally in the area of neural valuation.

The interest in valuation also led to a growing interest among neuroscientists in the role of emotion in neural processing, as emotions are typically regarded in neuroscience as reflective of reward processing, which in turn led to the growing appreciation of the role of emotions in decision-making. This was seen most strikingly in the neuroscience literature of the time in neurological patients whose impairment in emotional processing was accompanied by major changes in their decision-making and social interaction, which helped spawn a growing social neuroscience (Damasio 1994). As this work progressed, however, it became apparent to many neuroscientists that neuroscience lacked well-developed methods to probe reward processing, decision-making and social interaction. Hence, neuroscientists increasingly looked to economics, not only for quantitative models of decision-making under uncertainty but also for quantitative models of social interaction, particularly behavioural game theory (Sanfey *et al.* 2003; King-Casas *et al.* 2005).

The influx of neuroscientists into neuroeconomics is a major driving force behind its growth and may well represent the majority of researchers who regard themselves as engaged in neuroeconomic research. It is interesting to note that out of cognitive science grew two overlapping but somewhat distinct fields: from the top-down grew cognitive neuroscience while from the bottom-up grew computational neuroscience. The distinctions between these fields lie in differing research agendas, an ordering of major open questions, increasing specialization, along with sufficiency conditions on theory formation. It remains an open question what disciplinary form neuroeconomics will take, whether it will become an entity distinct from economics and neuroscience departments (as was the case with cognitive science), whether it will fragment into specializations within economics and neuroscience, or whether it will take some other form. In this regard, it is worth noting that the goals of neuroscientists within neuroeconomics are distinct from those of economists for a variety of reasons, which may order their research agenda in ways that diverge from their economic collaborators and may shape the ultimate form neuroeconomics takes. For example, whereas economists within neuroeconomics face the economic critique that their work is too reductive, from the perspective of neuroscience it may not be reductive enough, particularly work at the functional imaging level. For this reason, many neuroscientists are eager to make links with neurophysiology (some neuroeconomics labs currently employ both neurophysiology and fMRI) or other lower-level methods in an effort to uncover underlying neural mechanisms. A likely consequence of this is that the number of animal models of economic behaviour will grow. While some economic collaborators may embrace these developments, it may also sharpen the external criticism of neuroeconomics as it moves further away from traditional focuses. Relatedly, a research priority for many neuroscientists will be to apply neuroeconomic methods to clinical populations, such as subjects with bipolar disorder, schizophrenia, and autism. Indeed, the methods of neuroeconomics promise to shed important light on these illnesses and disorders, and are currently being utilized among these populations (Paulus 2007). These research priorities may not correspond to those of economists.

For these and other reasons, neuroscientists will likely be less concerned than their economics colleagues with the question, "is it good economics?" For neuroscientists, their interests will lie in whether the use of economic models and tools helps them make progress as measured by neuroscientific standards, whether neuroscientific journals publish their findings, and whether neuroscientific funding sources fund their research. They may be less moved by the concern that the research meets the expectations of economics, and this divergence may create impediments to interdisciplinary collaboration.

IS NEUROECONOMICS A SUCCESS?

The early days of cognitive neuroscience and computational neuroscience were, in retrospect, replete with non-starters, overly ambitious agendas, and unfulfilled promissory notes. It would, however, have been extremely premature to have judged the success or failure of either new discipline based on those early days. Likewise, it strikes me as extremely premature to judge neuroeconomics on the basis of its current results. Rather, the lessons of cognitive neuroscience and computational neuroscience, both of which are unquestionably extremely productive enterprises today, is that new interdisciplinary enterprises require a gestation period in which investigators learn the practices, standards and vocabulary of disciplines, and likely most importantly young investigators and graduate students are trained in both.

It is likely that these disciplinary differences account for Harrison's charge that the lack of publicly available fMRI data is a 'dark secret' of the field. Neuroscience, like many other disciplines such as physics, has not traditionally made raw experimental data publicly available or, for that matter, generally available to researchers within the field. My own speculation is that this tradition in neuroscience grew out of neurophysiology, in which a single experiment often took years of often heroic effort to complete, particularly if it involved primate behaviour. Rather, like the standard in most other experimental fields, replication was at the level of experimental parameters, not data-sharing. There have been, however, 'neuroinformatic' attempts to make raw fMRI data available, such as the fMRI Data Center (http://www.fmridc.org/f/fmridc). Whether these are actually useful datasets is, I think, an open question, as raw fMRI data typically requires minute details of the experimental parameters to be useful. There is also the additional complication that to release fMRI data would likely require a revision of standard informed consent protocols authorizing its release even in anonymized form, as the relationship between fMRI data and medical information continues to be a major issue (such as incidental findings).

THE ROLE OF FMRI IN NEUROECONOMICS

There is no doubt that functional magnetic resonance imaging has played a prominent role in the emergence of neuroeconomics. It is worth noting, however, that in a 1988 perspective on cognitive neuroscience, Churchland and Sejnowski's widely reproduced figure of the spatial and temporal resolution of the experimental techniques of cognitive neuroscience did not include fMRI (and has been subsequently updated) (Churchland and Sejnowski 1988). In other words, cognitive neuroscience emerged prior to fMRI and relied instead on computational links across levels of organization. fMRI has become a workhorse of cognitive neuroscience because it represents the best tradeoff in terms of spatial and temporal resolution of the (limited) non-invasive probes of human brain function. Prior to its development, there was virtually no non-invasive probe for studying the human brain (electroencephalogy was available, but had extremely poor spatial resolution). In the absence of such non-invasive human probes, neuroscience relied on animal model systems, which proved enormously successful for investigating such areas as the cellular basis of learning, but for many key human domains there are no corresponding animal models (e.g. language, strategic interaction involving theory of mind). While fMRI has rapidly developed, it nonetheless has basic limitations, including the temporal properties of the haemodynamic response, its relation to underlying physiological events, and limited spatial resolution. For these reasons, neuroscientists typically regard it as one to be utilized as a convergent technique alongside evidence from other neural sources, including data from animals models and computational accounts (whereas fMRI has generated a great deal of enthusiasm outside neuroscience, among many neuroscientists it is considered an extremely limited tool). Increasingly, as neuroeconomics research progresses, it will need to integrate these convergent methods alongside fMRI, as indeed is already occurring (O'Doherty *et al.* 2006).

IS NEUROECONOMICS A RADICAL NEW THEORY OR AN INCREMENTAL CHANGE?

Harrison's interesting discussion of what constitutes an economic agent in relation to the work of Don Ross (Ross 2005), and whether this notion remains unitary, incorporates a dual process theory, or something else raises an important set of issues. The explanatory vocabulary of cognitive neuroscience, and by extension, neuroeconomics, is couched firmly in the familiar language of "folk psychology", whereby human behaviour is explained in terms of intentional states, such as propositional attitudes, and (largely) semantically coherent transitions among those states (confirming to a principle of rationality). However, it was always the suggestion of cognitive neuroscience that a more complete theory may radically revise this framework. There are some strong indications that this framework is already under pressure, as traditional categories such as reason and emotion appear not to map onto neural systems in an unproblematic manner, and our functional interpretation of neural structures remains highly simplistic (Pessoa 2008). Further, it is important to bear in mind that while neuroscience has made remarkable advances, many of its most fundamental questions remain unanswered. To mention just a few, the information-processing capacity of single neurons, the nature of the neural code (how neurons code/decode information), how information is integrated within small neighbourhoods of neurons, to say nothing of long-range communication, remain fundamental questions. As progress is made on these fundamental issues within neuroscience, it is likely that they will necessitate large-scale revision of higher-level theories.

WHAT METHODOLOGICAL DIRECTION SHOULD NEUROECONOMICS TAKE?

Going back to Jevons and Pareto, although both sought to establish a very different foundation for economics, they both shared deep methodological commitments. In particular, both adopted the common methodological principles of their day, John Stuart Mill's concrete deductive method, whereby from a small set of obvious laws (Jevons's example is that a greater gain is preferred to a smaller one) one may deduce complex economic phenomena. These methodological commitments are clearly in evidence throughout the axiomatic approach of neoclassical economics.

Neuroeconomics clearly requires a distinct methodology as an experimental discipline and one in which there as yet appears to be few organizing principles around which neuroscience can be organized. Harrison, based on a discussion of (Glimcher 2004), considers the possibility that neuroeconomics adopt the top-down methodology David Marr (1982) developed, whereby a problem is decomposed along three distinct levels at which an information processing system must be understood, the computational, the algorithmic, and the implementational. These three levels also correspond to a methodology, whereby a computational level precedes the algorithmic, and so on. Marr's methodology, however, grew out of many of the core commitments of cognitive science (indeed, Marr cites Chomsky's (1959) theory of universal grammar as paradigmatic of the computational level). In retrospect, Marr's proposed methodology turned out to be a failure even within his own area of vision (Churchland *et al.* 1994). From the computational level, it seemed obvious that the computational goal of vision is to provide an accurate internal representation of the external world. This, combined with Marr's modular decomposition of a task into discrete sub-tasks (combined with Fodor's claim that modules operate in a bottom-up fashion), led Marr to the notion that the goal of vision was to produce what he called a 2 $^1/_2$ - D sketch through a hierarchically organized series of computations decomposed into sub-problems (such as shape from shading, motion, etc.), that were subsequently integrated, and then interfaced with visual cognition. This not only mistook the ecological goals of vision (whereby some visual circuits bypass cortical centres altogether and project directly to spinal cord indicating the goal of vision is not simply to replicate the visual world, but to link visual information with adaptive action such as predator evasion), but also neglected the non-hierarchical nature of visual information-processing, and the non-modular nature of cortical information-processing (Sporns and Zwi 2004). As a matter of discovery, then, a detailed understanding of the underlying physical implementation is necessary for theory construction in neuroeconomics. For these reasons, neuroeconomists should be pragmatic in the use of available data from animal models, information regarding neural pathways and projections, computational models of neural function, functional imaging, and behavioural data. As cognitive science illustrates, the major challenge facing theory formation in the neural and behavioural sciences is that of being under-constrained by data, and it is precisely in the capacity of cognitive neuroscience to provide convergent evidence across multiple levels of organization that neuroeconomics has the most promising prospects.

REFERENCES

Baker, L. 1989. Instrumental intentionality. *Philosophy of Science* 56: 303–16.

Bruni, L. and R. Sugden 2007. The road not taken: how psychology was removed from economics, and how it might be brought back. *The Economic Journal* 117: 146–73.

Chomsky, N. 1959. A review of B. F. Skinner's verbal behavior. *Language* 35: 26–58.

Churchland, P. S., V. S. Ramachandran and T. J. Sejnowski 1994. A critique of pure vision. In *Large-Scale Neuronal Theories of the Brain*, ed. C. Koch and J. L. Davis, 23–51. Cambridge, MA: MIT Press.

Churchland, P. S. and T. J. Sejnowski 1988. Perspectives on cognitive neuroscience. *Science* 242: 741–5.

Damasio, A. R. 1994. *Descartes' error: emotion, reason, and the human brain*. New York: G.P. Putnam.

Dayan, P. and L. F. Abbott. 2001. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

Dennett, D. 1987. *The intentional stance*. Cambridge, MA: MIT Press.

Friedman, M. 1953. The methodology of positive economics. In *Essays in Positive Economics*, 3–43. Chicago: University of Chicago Press.

Gazzaniga, M. S. 2004. *The cognitive neurosciences*, 3rd Edn. Cambridge, MA: MIT Press.

Glimcher, P. W. 2004. *Decisions, uncertainty, and the brain: the science of neuroeconomics*. Cambridge, MA: MIT Press.

Gul, F. and W. Pesendorfer 2008. The case for mindless economics. In *The handbook of economic methodologies*, ed. A. Caplin and A. Schotter. Oxford: Oxford University Press.

Hubel, D. H. and T. N. Wiesel 1959. Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology* 148: 574–91.

Jevons, W. S. 1871. *The theory of political economy*. New York: MacMillan and Co.

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

Koob, G. F. and N. R. Swerdlow 1988. The functional output of the mesolimbic dopamine system. *Annals of the New York Academy of Sciences* 537: 216–27.

Kuffler, S. 1953. Discharge patterns and functional organization of the mammalian retina. *Journal of Neurophysiology* 16: 37–68.

Lagueux, M. 1994. Friedman instrumentalism and constructive empiricism in economics. *Theory and Decision* 37: 147–74.

Maas, H. 2005. *William Stanley Jevons and the making of modern economics*. New York: Cambridge University Press.

Markowitz, H. 1952. Portfolio selection. *Journal of Finance* 7: 77–91.
Marr, D. 1982. *Vision : a computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.
Minsky, M. and S. Papert 1969. *Perceptrons: An introduction to computational geometry*. Cambridge, MA: MIT Press.
Montague, P. R., P. Dayan, C. Person and T. J. Sejnowski. 1995. Bee foraging in uncertain environments using predictive hebbian learning. *Nature* 377: 725–8.
Montague, P. R., P. Dayan and T. J. Sejnowski. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience* 16: 1936–47.
O'Doherty, J. P., T. W. Buchanan, B. Seymour and R. J. Dolan. 2006. Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. *Neuron* 49: 157–66.
Paulus, M. P. 2007. Decision-making dysfunctions in psychiatry – altered homeostatic processing? *Science* 318: 602–6.
Pessoa, L. 2008. On the relationship between emotion and cognition. *Nature Reviews Neuroscience* 9: 148–58.
Preuschoff, K., P. Bossaerts and S. R. Quartz. 2006. Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* 51: 381–90.
Putnam, H. 1975. *Mind, language, and reality: Philosophical papers, Vol.* 2. Cambridge: Cambridge University Press.
Pylyshyn, Z. 1984. *Computation and cognition*. Cambridge, MA: MIT Press.
Quartz, S. R. and T. J. Sejnowski. 1997. The neural basis of cognitive development: a constructivist manifesto. *Behavioral and Brain Sciences* 20: 537–56; discussion 556.
Rogers, T. T. and J. L. McClelland. 2004. *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
Rosenblatt, F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65: 386–408.
Ross, D. 2005. *Economic theory and cognitive science: microexplanation*. Cambridge, MA: MIT Press.
Rumelhart, D. and J. L. McClelland. 1986. *Parallel distributed processing: explorations in the microstructure of cognition*, 2 Vols. *Computational models of cognition and perception*. Cambridge, MA: MIT Press.
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.
Sporns, O. and J. D. Zwi 2004. The small world of the cerebral cortex. *Neuroinformatics* 2: 145–62.
Sutton, R. S. and A. G. Barto. 1998. *Reinforcement learning: an introduction, adaptive computation and machine learning*. Cambridge, MA: MIT Press.
Werbos, P. 1994. *The roots of backpropagation*. New York: Wiley.