Cognitive variables and parameters in economic models

(forthcoming in R. Sun, ed., *Grounding Social Sciences in Cognitive Science*, MIT Press)

Don Ross

School of Economics
University of Cape Town

Center for Economic Analysis of Risk
Georgia State University

don.ross@uct.ac.za

Abstract

The paper critically reviews the impact of models of cognitive architecture on economics. A historical survey is provided that indicates the reasons for most economists' non-participation, as either producers or consumers, in the cognitive revolution. The remainder of the paper shows, partly by means of a case study of economists' models of temporally inconsistent consumption, that economists continue to bypass cognitive architectural considerations, even when modeling a phenomenon that they recognize as implicating representational and framing idiosyncrasies. This is not attributed to economists' alleged disinterest in empirical facts or over-commitment to normative rationality. Rather, it stems from the fact that, on the one hand, most behavioral economists and neuroeconomists remain strongly committed to methodological individualism; while, on the other hand, experimental economists who follow Vernon Smith in emphasizing ecological rationality are so far mainly interested in aggregate responses. This divergence of assumptions and interests results in neglect among economists of Marr's 'algorithmic' level of cognitive task analysis – precisely the ground on which people's strategic problems might be considered in tandem with the biological restrictions on the computations they perform.

JEL classification: B21, B25, B41, D01, D03, D87

1. Introduction: overview of cognitive variables in economics

The current relationship between cognitive science and economics is complex, contested and unsettled. An overview of it therefore cannot be a summary of points of consensus, but must indicate and explain points of disagreement. The task is complicated by the fact that the economics / cognitive science relationship inherits residues from a long history of interanimations between economics and psychology that have implicitly shaped the attitudes and methodological biases of practitioners. In consequence, particular economists' pronouncements of their views on the role of cognitive science are often less informative than what is revealed by their practices. Finally, at all times in the history of economics, an orthodoxy or mainstream methodology has coexisted with various dissenting heterodoxies. In this chapter, when it is said that 'economists' of a given time and place thought or think such-and-such, this should be understood as referring to the orthodoxy of that time and place.

Let us begin with a broad sketch of the evolution of economists' attitudes toward psychology prior to the coming of the cognitive revolution in the 1960s. The earliest major thinkers on topics that preoccupy modern economists, in particular Smith and Hume, did not concern themselves with a boundary between economics and psychology because no traditions or institutions had yet stabilized such a frontier. Read from a time in which institutionalized disciplinary borders are being re-drawn, Smith and Hume often seem more modern and sensible than their successors. However, as there is not space in this chapter to do any justice to the rich recent scholarship on Smith's thought, it will be set to one side, and the sketch will open with 1870, the beginning of the decade in which the early marginalists self-consciously separated economics from political philosophy.

The founders of modern economics, Walras, Jevons and Menger, disagreed on various matters. However, all shared the view, doubted by some of their successors, that aggregate economic phenomena and regularities should be explained by reference to the actions and dispositions of individual people. That is, all were what we would now call methodological individualists

Jevons (1871) exemplified what we might call 'the English philosophy', explicitly derived from Bentham's utilitarian psychology. According to this perspective, psychological mechanisms that are partly innate and partly adaptive endow people with natural relative subjective valuations of experiences. These valuations are then directly reflected in their choices amongst alternative consumption baskets. Two key general properties of the natural human psychology of valuation according to Jevons were (1) that wants are insatiable, so that the person who has fulfilled her basic needs will devote her energies to the pursuit of more complex consumption experiences; and (2) that people attach decreasing marginal value to further increments of any given stream of consumption goods as their stock and / or their rate of flow of that stream increases. These principles lead directly to emphasis on *opportunity cost* – analysis of the value of a choice in terms of the alternative possibilities it forecloses – that has characterized all subsequent mainstream economics.

Jevons was optimistic that empirical research in psychology would uncover the mechanisms that give rise to the principles of non-satiation and decreasing marginal utility, and also to the sources of variance in individual tastes. It is therefore appropriate to say that he anticipated eventual unification of economics with psychology; the need to drive economic analysis from the top-down logic of maximization of utility through minimization of opportunity cost was, on his view, a temporary artifact of ignorance of psychological mechanisms. One important caveat is necessary in this regard. Along with many of the first generation of modern economists, Jevons doubted that people's social and moral preferences are determined by mechanisms that empirical psychology would reveal. Such 'higher wants' were taken by him, to be functions of distinctive moral reasoning that people must learn from teaching and philosophical reflection. Thus Jevons and many of his immediate successors thought that economists should concentrate their attention on selfish consumption, not because they believed that people are narrowly selfish by nature, but because they thought that other-regarding behavior remained outside the scope of both economics and mechanistic psychology.

This historical English view of the relationship between economics and psychology contrasts with a rival Austrian one, articulated originally by Menger (1871). Its mature exposition was given by von Mises (1949), and the boiled-down articulation and simplification through which most economists know it was provided by Robbins (1935). According to this view the foundations of economics lie in broad psychological truisms that are discovered by introspection and logical reflection, and which are not susceptible to significant modification by discoveries of scientific psychology concerning their underlying basis. (Jevons's two principles above were held by Robbins to be among the truisms in question.) Robbins explicitly argued that economists in their professional capacity should ignore psychology. Though the psychology he had in mind, given the time at which he wrote, was behaviorist, his premises would apply equally to contemporary cognitive science. In addition, Robbins denied that economic analysis applies only to choices amongst 'lower', material, goods, arguing that opportunity cost influences the consumption of anything that is scarce, including sources of social, moral and intellectual satisfaction.

The methodological views of the Austrians have enjoyed more success among economists than their first-order economic doctrines. The period from end of the nineteenth century to the 1980s saw a steady retreat of the English view. In consequence, economists, with the notable exception of Herbert Simon, were not important contributors to, or an important audience for, the early stages of the cognitive revolution. However, the past four decades have witnessed an increasing resurgence of Jevons's position as so-called behavioural economists and their collaborators from psychology have empirically unearthed cognitive biases and framing effects that influence economic decisions (Kahneman *et al* 1982), posited computationally fast and frugal heuristics that allow people to exploit environmental regularities and ignore information that would otherwise overwhelm limited processing capacities (Gigerenzer *et al* 1999), and modeled endogenous, dynamic construction of preferences in response to experience (Litchtenstein & Slovic 2006).

Behavioral economists regularly accuse their less psychologically informed colleagues of applying false models of economic agency, derived from normative attributions of rationality rather than from empirically tested hypotheses about the actual cognitive resources with which evolution has equipped our species, and which learning and development can refine only up to limits. (For a widely cited example of such criticism, see Camerer *et al* [2005]. Angner & Loewenstein [forthcoming] is more comprehensive in coverage and somewhat more cautious in tone. Ariely [2008] is one of many recent popular announcements of revolution against what is held to be orthodox dogma.) Ross (2008) sorts the discoveries that are held to challenge conventional economics into four sets: (1) findings that people don't reason about uncertainty in accordance with sound statistical and other inductive principles; (2) findings that people behave inconsistently from one choice problem to another as a result of various kinds of framing influences; (3) findings that people systematically reverse preferences over time because they discount the future hyperbolically instead of exponentially; and (4) findings that people don't act so as to optimize their personal expected utility, but are heavily influenced by their beliefs about the prospective utility of other people, and by relations between other peoples' utility and their own. Note that findings in the set (4) challenge mainstream economics only insofar as it is taken to be committed to the hypothesis that individual economic behavior is motivated by narrow selfishness; but this attribution applies only to rhetoric of some well known economists, rather than to any proper element of economic theory (Ross 2005, Binmore 2009). Findings in sets (1) – (3) challenge the idea that people's economic behavior conforms to normative ideals of rationality. It is natural to suppose that this might result from features of their brains and/or cognitive architectures that reflect adaptive sculpting by particular environments. This is why results of behavioral economics are taken to imply the inclusion of cognitive processing variables in economic models.

The past decade has in addition seen the rise of the interdisciplinary field of neuroeconomics, which studies the mediation of choices and valuation of consumption alternatives by neural computations in specific brain areas (Glimcher *et al* 2009). Standard methodologies in neuroeconomics involve confronting monkeys and humans with reward choices while measuring their neural activity through, respectively, invasive single-cell recording and functional Magnetic Resonance Imaging (fMRI). Neuroeconomists often interpret their work as realizing Jevons's anticipation of a time when psychologists would directly observe mechanisms responsible for varying subjective judgments of value.

Some aspects of Jevons's view have not enjoyed revival. No behavioral economists follow him in thinking that social motivations lie outside their domain. If anything, they tend to invert the Jevonsian picture: narrowly selfish conduct is typically held to be a learned, cognitive response to impersonal market institutions, while more primitive, automatic and cognitively encapsulated neural-computational mechanisms incline people to sympathetic action for the benefit of perceived affiliates, and inherent anti-sympathy for and distrust of non-affiliates.

The side of opinion descended from the Austrian view has also evolved. In particular, the attempt to ground economic generalizations in introspectively accessible psychological truisms has been largely abandoned. Completing an effort begun by Pareto (1909), Samuelson's (1938, 1947) Revealed Preference Theory (RPT) derived the existence of sets of preferences mappable onto the real numbers by monotonic, complete, acyclical, and convex functions from observable schedules of aggregate demand. Samuelson would have preferred not to call these 'utility' functions, but the lure of semantic continuity turned out to be decisive. Notwithstanding the semantic suggestion of 'revealed' preference, his utility functions were intended as descriptions of actual and hypothetical behavior, not indicators of hidden inner states. It is common to attribute the motivation for this to the behaviorism and positivism that dominated the psychology and social science of the 1930s, 40s and 50s, and certainly this influence played its part. However, the main impetus was the strongly felt doubt, clearly articulated by Robbins and also expressed by Pareto, Fisher and Hicks, that economic generalizations about aggregate responses to changes in supply, demand or prices are hostage to contingencies about individuals that are discoverable by scientific psychology. Post-Samuelson economists began from observable aggregate demand. The central result of RPT was that if this has certain testable properties then the existence of continuous preference fields is implied. The causal influences that stabilize such fields might or might not be properties of individual psychologies; the revealed preference theorist disavows professional interest in this question. Becker (1962) showed that the fundamental property of the standard model of the market – downward sloping demand for any good given constant real income – depends on no claim about the *computational* rationality of any agent; it depends only on the assumption that households with smaller budgets, and therefore smaller opportunity sets, consume less.

The dominance of RPT in economics was the primary reason for the initial disinterest shown by economists, outside of a small dissident minority championed by Simon (1969, 1972), in the cognitive revolution. Two main developments that unfolded across economics during the 1970s undermined this cross-disciplinary isolationism.

The first of these developments was the extension of game theory (GT) to ever-widening subfields of microeconomics. Resistance to incorporating individual psychological variables in economic models was based to some extent on practical concerns about mathematical tractability: the constrained optimization methods used by economists to solve most problems did not readily allow for representation of idiosyncratic preference structures among different agents, of the sort that psychologically derived hypotheses about motivation tend to furnish. This was precisely the restriction lifted by GT. Furthermore, GT didn't merely make the introduction of psychological variables into economic models *possible*; in specific ways it *invited* such introduction. Application of GT in economics required the enrichment of utility theory that von Neumann and Morgenstern (1944) and Savage (1954) provided in order to incorporate players' uncertainty about the valuations of and information available to other players. This enrichment was elucidated at every step by formalization of concepts drawn from everyday psychology. Thus the non-mathematical version of the vocabulary of GT came to be full of psychological notions, particularly beliefs, conjectures, and expectations. GT *can* be given a strictly behaviorist interpretation, according to which one uses it to guide inferences about players' stable behavioral orientations through observing which vectors of possible behavioral sequences in strategic interanimation are Nash equilibria. But the power of such inferences is often limited because most games have multiple Nash equilibria. Efforts to derive stronger predictions led a majority of economic game theorists in the 1980s to follow the lead of Harsanyi (1967) in interpreting games as descriptions of players' *beliefs* instead of their *actions*. On this interpretation, a solution to a game is one in which all players' conjectures about one another's preferences and (conditional) expectations are mutually consistent. Such solutions are, in general, stronger than Nash equilibria, and hence more restrictive. The resulting 'equilibrium refinement program' drew game theorists deep into evaluation of alternative, speculative computational models. This program has somewhat petered out over recent years after devolving into unproductive philosophical disputes over the meaning of 'rationality'; however, along the way it familiarized many economists who had ignored Simon's work with issues around computational tractability.

Second, the formal completion of general equilibrium theory by Arrow and Debreu (1954) required clearer formalization of the idea of the economic agent (Debreu 1959). In particular, it was necessary to assume that the agents in general equilibrium models can rank all possible states of the world with respect to value, and that they never change their minds about these rankings. As argued by Ross (2005), nothing in this regimentation requires that agents be interpreted as perfectly coextensive with natural persons. However, a crucial intended point of general equilibrium theory since its inception had been to serve as a framework for thinking about the consequences of changes in exogenous variables, especially policy variables, for welfare. Improvements and declines in the feelings of natural persons about their well being is what most people, as a matter of fact, mainly care about. Thus if the loci of preference fields in general equilibrium theory are not at least idealizations of natural persons, then it is not evident why efficiency, the touchstone of general equilibrium analysis, should be important enough to *warrant* touchstone status. Once one acknowledges pressure to include models of distinct human consumers in general equilibria, however, troubles arise. The so-called excess demand literature of the 1970s (Sonnenschein 1972, 1973; Mantel 1974, 1976; Debreu 1974) showed that although all general equilibria are efficient, there is no unique one-to-one mapping between a given general equilibrium and a vector of individual demand functions. (Put more plainly, for a given set of demand functions there is more than one vector of prices at which all demand is satisfied.) In tandem with the Lipsey-Lancaster (1956) theory of the second-best, the Sonnenschein-Mantel-Debreu Theorem challenged the cogency of attempts by welfare economists to justify policy by reference to merely inferred (as opposed to separately and empirically observed) subjective preferences of consumers. Note that this problem arises whether one assumes an atomistic or an intersubjective theory of the basis of value. Nevertheless, the excess demand results shook the general postwar confidence that if one attended properly to the aggregate scale then specific preferences of individuals could be safely ignored. Thus we now find some leading economists (e.g., Bernheim & Rangel 2008) devoting effort to incorporating psychological limitations and agent heterogeneity into normative models.

These developments have certainly not convinced all economists to join a 'back to Jevons' movement. Gul and Pesendorfer (2008) recently published a manifesto against such incorporation. They deny as a matter of principle that anything can be learned about economics by studying neural information processing or cognitive structures. According to Gul and Pesendorfer, the task of positive economics is to predict choices as functions of shifts in incentives and opportunity sets. Such choices are abstract constructs rather than directly observable phenomena, as are all relationships into which they enter. Though Gul and Pesendorfer concede that the mechanics of choice presumably involve neural computations, they see this as no more relevant to the *economics* of choice than is the fact that every choice must occur at some specific set of geographical coordinates. The intended value of such extreme abstraction lies in achieving generalizations of maximum scope. Economic models, on this conception, should be represented by reduced-form equations that leave psychological and neural processing variables inside black boxes.

To judge from responses to Gul and Pesendorfer's paper that have appeared to date, few economists endorse such strong disciplinary isolation. Harrison (2008), despite being sympathetic to the idea that economic models should include no more psychological variables than are necessary to best fit aggregate empirical choice data in large subject samples, argues that the idea that economists could black-box all processing variables in reduced-form equations is a theorist's illusion that could not be seriously entertained by any researcher with experience in modeling real experimental data from human subjects. However, the Gul and Pesendorfer critique is useful for establishing a general methodological standard by which to evaluate claims about the relevance of cognitive influences on economic relationships. The general goal of the economist is to propose and test models that feature patterns of choices as dependent variables, with incentives and budgets among the independent variables. Which additional independent variables, including variables for cognitive structures and processes, should be included in empirically adequate and predictively powerful generalizations is a matter to be determined by evidence, not a priori reflection. Most economists, including many who conduct experiments, probably agree with Gul and Pesendorfer that, in light of the value of parsimony in modeling, the burden of argument in any given case lies with the theorist who proposes additional variables or parameters.

This is not the view promoted by 'radical' behavioral economists who argue that the discipline stands in need of a 'paradigm shift' following a long history of inadequate psychological realism. Popular accounts, and occasional scholarly publications (e.g. Zak 2008), suggest that the need to condition economic models on cognitive and/or neural structures and processes follows directly from the fact that economic behavior must in every instance be produced by cognitive and/or neural computation. The invalidity of this general form of reasoning (which would license the conclusion that variables from fundamental physics must appear in every generalization of every science) should be obvious. Nevertheless, someone might reasonably wonder how there could be generalizations about people's responses to changes in prices or other incentives that are *not* based on generalizations about human cognitive architecture.

A partial answer to this question may lie in what Vernon Smith (2007) refers to as the *ecological* nature of economic rationality. Ecological rationality emphasizes the extent to which people at least sometimes approximate consistent, optimizing rationality in their choice behavior by means not of computational marvels they achieve with their raw cognitive or neural apparatus, but thanks to what Hutchins (1995) and Clark (1997) call "scaffolding." Social scaffolding consists of external structures in the environment that encode culturally accumulated information and constrain and channel behavior through processes that Grassé (1959) called *stigmergic*. Grassé's example was of termites coordinating their colony building actions by means of algorithms that take modifications to the environment made by other termites as inputs; the modifications themselves, in the contemporary terminology, are scaffolding. The social scaffolding elements that support stigmergy can become *cognitive* scaffolding when agents internalize them to cue thoughts through association. Sun (2006, p. 13) summarizes the idea as follows:

> First of all, an agent has to deal with social and physical environments. Hence, its thinking is structured and constrained by its environments … Second, the structures and regularities of an environment may be internalized by agents, in the effort to exploit such structures and regularities to facilitate the attainment of needs. Third, an environment itself may be utilized as part of the thinking (cognition) of an agent …, and therefore it may be heavily reflected in the cognitive process of an agent.

Consider a simple example of social scaffolding first. Most people who regularly consume alcohol avoid becoming addicted to it. Some may do this by carefully choosing consumption schedules that allow them to maintain high equilibrium levels of the neurotransmitters serotonin and GABA, which inhibit control of consumption by midbrain dopamine systems that are vulnerable to obsessive recruitment of attention by short reward cycles (Ross *et al* 2008). This typically involves manufacture and maintenance of personal rules that, as Ainslie (2001) demonstrates, require complex cognitive self-manipulation, at least until they become habitual. Another person might avoid addiction without calling upon any effortful willpower by becoming an airline pilot, thereby facing extreme sanctions for indulgence that are so strong as to prevent temptation from ever arising (Heyman 2009). The first person thus relies upon inboard cognition to regulate consumption, while the second person is restricted by social scaffolding. To an economist persuaded by the Gul and Pesendorfer view, what is relevant about the two cases is captured by a single model of that portion of the population that does not drink more when the relative price of alcohol falls, and the cognitively interesting dynamics of the first case are left aside.

As an example of cognitive scaffolding, consider a person whose assets consistently increase, not because she knows how to gather information or think about financial markets, but because she has learned the simple rule “buy through whichever broker was used by the friend who suggests an asset purchase”, and she has lots of friends who tell her about their investments. Her simple heuristic causes her to distribute her business across many brokers, and thus she hedges.

The phrase ‘scaffolding’ has not yet entered the economics literature. However, economists often speak of ‘institutions’ in a way that is general enough to incorporate the idea. An example that is prominent in the literature is Gode & Sunder’s (1993) simulation of ‘zero intelligence’ agents that participate in a double auction experiment subject to budget constraints and very simple rules. These rules are that sellers do not charge more than their marginal cost, and buyers do not make negative offers or offers above a fixed ceiling. Otherwise the agents bid randomly. The efficiency of these simulated markets matches that achieved by human subjects. This suggests that efficiency of outcomes in such markets may result from the ecological rationality of the institutional rules, rather than sophisticated ‘inboard’ computations. Sunder (2003) discusses generalizations of this result, and further economic applications.

Widespread use of scaffolding to achieve economic rationality does not suggest that cognitive variables are irrelevant to economic modeling; quite the contrary. People making decisions by use of fast and frugal heuristics will tend to suffer catastrophic performance collapses if their environments change drastically, and this will only be successfully predicted by a model that captures the cognitive structures underlying their choices. What the importance of scaffolding *does* indicate is a main reason for the invalidity of direct inferences from the stability of an economic regularity to the existence of any matching cognitive regularity. In the case of any given regularity, institutions may be carrying more or less of the load. The issue is always purely empirical and contingent.

There is an orthogonal sense in which appeal by economists to institutional scaffolds implies indirect reference to cognitive elements. Institutions are not merely self-reproducing social regularities; they are in addition partly sustained by norms. Conte & Castelfranchi (1995) argue that norms are, in turn, more than just the focal points for coordination recognized (though still not formally axiomatized) by game theorists. Norms must be represented by agents *as* norms in order for agents who internalize them to expect to suffer the distinctive costs that accompany norm violation. Thus norms imply cognitive architectures that can support the relevant representations. Note, however, that this does not necessarily imply the inclusion of the implicated architectural variables in economic models. A model might simply distinguish between normatively regulated and other behavior, in ways that are econometrically detectible, without specifying the cognitive elements that must underlie the normative regulation; these could in principle remain black boxed. This strategy is exemplified, for example, by the well-known work of Bicchieri (2005), who incorporates norms into individual utility functions for application in game-theoretic models.

Over the past few decades, some economists have joined other social scientists in becoming interested in multi-agent models aimed at simulating social dynamics. Representative selections of this work can be found in Anderson *et al* (1988), Arthur *et al* (1997), and Blume & Durlauf (2005), all based on work emanating from the Santa Fe Institute in New Mexico. Its theoretical basis lies in evolutionary game theory, for which foundational mathematical relationships with classical equilibrium solution concepts have been extensively though incompletely studied (Weibull 1995, Cressman 2003). Almost all of this work to date involves cognitively simple agents, and thereby black boxes both the influence of cognitive variables on social interactions, and feedback effects of the latter on the former. The program urged by Sun (2006) for combining cognitive with social modeling has thus not yet been taken up by economists, except in a few isolated instances focused on restricted topics (e.g., Conte & Paolucci 2002). This might be regarded as surprising because, as will be discussed in detail in the next section, economists have devoted considerable attention to models of individual people as disunified agents. The explanation probably lies in the strong interdisciplinary preference in economics for models that can be represented by closed-form equations. There is little or no stability of state variables from one complex-system model of the economy to the next. As long as this is the case, economists are likely to doubt that this methodology promises to deliver accumulation of theoretical knowledge. Under such circumstances, it is perhaps less puzzling that few have ventured to combine complexity at the social scale with yet more complexity at the intra-agent scale.

At this point, therefore, there is little to be reported on the possibility that alternative models of cognitive architecture might support systematic differences in economic outcomes and dynamics. On the other hand, we can view the recent proliferation of economic models of cooperation and competition among intra-personal agents as preliminary steps toward the introduction of computational architectures that will allow for simulation of processes of mediation between microeconomics and macroeconomics along the lines suggested fifteen years ago by Conte & Castelfranchi (1995). In the next section, we will therefore review the current state of theoretical modeling of the disunified economic agent at a finer level of detail than has characterized the discussion so far. To provide additional focus, the comparison of models will concentrate on a specific target phenomenon, intertemporal discounting of utility as the basis for regretted consumption, procrastination and addiction. The investigation has a specific objective. Models of cognitive architecture, as opposed to models from cognitive neuroscience, have so far had little impact on or role in economic modeling. The history of efforts to model the sources of preference inconsistency shed some revealing light on this.

2. Theories of intertemporal discounting and impulsive consumption in psychologically complex economic agents

In economics, rationality is generally understood as comprising two elements: consistency of choice, and full use of such information as isn't too costly to be worth gathering. Behavioural economists typically motivate inclusion of psychological variables in economic models to the extent that (i) people are systematically inconsistent in their choices and (ii) are prone to act on scanty information even when more could be obtained with relatively little effort.

It is a common observation that people are prone to impulsive consumption, that is, patterns of choice that are subsequently regretted. Many people knowingly choose to consume substances they know to be addictive, then pay later costs to try to overcome their dependence on these substances. Others accumulate debts to credit card companies that they know cannot be rationalized by comparison with their levels of income and wealth. In almost all rich countries, the majority of retired people eventually regret their prior allocations of income as between savings and consumption. Ross (2010) argues that all of these phenomena can be understood as instances of procrastination, that is, as cases in which people postpone expenditure of effort aimed at optimizing their wealth in favor of enjoying their current income. In the case of addiction, procrastination involves a particularly interesting phenomenological twist. Being intoxicated is typically incompatible with effortful investment. Therefore, the person who procrastinates by getting high *commits* to her choice for a period of time, during which she is relieved of the anxiety of knowing she could 'get back to work' at any moment given some level of effort. Anxiety caused by a behavioral choice is a cost associated with that behavior, so people can be expected to be attracted to mechanisms that reduce this cost. This is part of the pleasure of intoxication. The other main source of its pleasure lies in the direct action of neurochemical mechanisms that stimulate pleasure centers and, more importantly, heighten stimulus salience by producing dopamine surges in the midbrain reward circuit (Ross *et al* 2008). A property of substance addiction that is common to procrastination processes in general is that the longer a person puts off her investments in the future, the higher becomes the cost of switching from procrastination to investment at any particular moment. Thus procrastination generally, and addiction specifically, have the form of 'behavioral traps'.

It is in principle possible to theoretically model procrastination and addiction in a way that is consistent with the hypothesis that a person is a unified rational agent. In general it is rational to discount future consumption relative to present consumption, simply because uncertainty increases with distance into the future; reward prospects may disappear, or the consumer may die or become incapacitated. Because people have idiosyncratic levels of tolerance for risk, rationality alone recommends no specific rate of intertemporal reward discounting. Under the idealization of a linear relationship between uncertainty and the passage of time, a rational agent should discount according to the formula

$$v_i = A_i e^{-kD_i} \qquad (1)$$

where v_i, A_i, and D_i represent, respectively, the present value of a delayed reward, the amount of a delayed reward, and the delay of the reward; *e* is the base of the natural logarithms; and the parameter $0 > k > 1$ is a constant that represents the influence of uncertainty and the agent's idiosyncratic attitude to risk. (Note that *k* is a psychological variable.) Becker & Murphy (1988) showed that there are values for these variables and parameters such that the agent will rationally choose to consume in a way that causes her welfare to steadily decline *over* time, because *at* any given moment of choice she is best off choosing to consume an addictive drug or other good that improves the payoff from future procrastination relative to future prudence while lowering the value of both relative to what would have been available had prudence been chosen in the first place.

This so-called 'rational addiction' model, using either an average value of *k* or some motivated distribution of *k*-values, is widely applied by economists engaged in such enterprises as discovering tax rates on addictive goods that are optimal with respect to revenue and / or public health. However, most economists now regard the model as inadequate for purposes of representing and predicting the consumption pattern of a specific individual agent. Addicts, in particular, seldom exhibit life-cycle consumption patterns that accord with it. According to the model, *if* an addict successfully overcomes withdrawal *and* intended at the time of entering withdrawal to quit for good, then either she should never relapse or she should never permanently quit. But the overwhelming majority of addicts repeatedly undergo the rigors of withdrawal, planning never to relapse, and then return to addictive consumption – before eventually quitting successfully. Thus the rational addiction model is empirically refuted by a basic aspect of the standard addictive pattern. (For the detailed logic underlying this argument, see Ross *et al* 2008, Chapter 3.)

The roots of this misprediction by the rational addiction model lie in the fact that it has no device for allowing an individual's preferences to change over time while preserving her identity as the same agent. The economist trying to cope with this problem can do so by one of three strategies for trading off the two aspects of rational agency (stability of preference and full use of information) against one another. Her first general option is to suppose that people procrastinate (consume impulsively) in full awareness that they are doing so, but exhibit intertemporal preference inconsistency. That is, they choose courses of behavior that are rationalized by the payoff that would accrue to completion of various specific investments, then choose not to complete the investments in question, then subsequently reverse preferences a second time and suffer regret over their irresoluteness. This strategy *diachronically* divides the agent into a *sequence* of sub-agents. A second general possibility is to model procrastinators as having intertemporally consistent utility functions but as lacking accurate information about their probable future behavior when they choose schedules of activities. Since what is mainly of interest is *recurrent* procrastination, the poverty of sound expectations here must be comparatively radical: the procrastinator fails to learn to predict future procrastination from her own history of past procrastination. This approach further divides into two strategies. One is to *synchronically* divide the agent into a *community* of sub-agents, in which different sub-agents have different information, and communication among them is imperfect. The alternative is to preserve agent unity by 'shrinking' the agent and moving aspects of cognitive architecture or of the brain onto the environment side of the agent / environment boundary. In this framework parts of the agent's mind or brain that are not 'parts' of the agent generate exogenous and unpredicted costs and benefits for the agent at particular points in time.

This choice is not the only dimension of modeling discretion concerning procrastination / impulsive consumption. The modeler must also decide which discounting function to use for fitting the intertemporal choice behavior of the whole agent. Musau (2009) surveys the full range of mathematically distinct options, but it is standard at the level of conceptual description to sort them into two families, hyperbolic functions and quasi-hyperbolic or 'β-δ' functions. The most common version of the hyperbolic function is Mazur's formula (2):

$$v_i = \frac{A_i}{1 + kD_i} \tag{2}$$

This produces discount curves that decline steeply from an origin at the temporal point of choice, then become (much) flatter between points further into the future. The quasi-hyperbolic function is a step function, borrowed from Phelps's & Pollack's (1968) model of the motives in intergenerational wealth transfers, and often referred to as 'β-δ'. This class of functions is expressed by

$$v_i = A_i\beta\delta^D \tag{3}$$

where β is a constant discount factor for all delayed rewards and δ is a per-period exponential discount factor. Where $\beta = 1$ the equation reduces to standard exponential discounting as in equation (1). Where $\beta < 1$, discounting is initially steeper up to some inflection point, then flattens. β-δ discounting predicts that value drops precipitously from no delay to a one-period delay, but then declines more gradually, and exponentially over all periods thereafter. In early applications to intertemporal inconsistency (e.g. Laibson 1994, 1997), β-δ discounting is defended as merely an idealized approximation to hyperbolic discounting, to be preferred for the sake of mathematical tractability (compatability with standard identifications of unique optima). However, more recently β-δ discounting has been promoted over hyperbolic discounting because the former lends itself to a particular interpretation in terms of cognitive neuroscience.

To explain this last point we must first introduce the third dimension of discretion in modeling intertemporally inconsistent choice. This is whether one intends a *molar* or a *molecular* interpretation of the model in respect of cognitive architecture and / or neuroscience.

The contrast between 'molar' and 'molecular' scales of description and explanation is a well established one in psychology, and was carried over into cognitive science as the 'bottom-up / top-down' distinction. Molar-scale descriptions situate behavioral systems in environmental contexts, sorting their dispositions and properties by reference to equivalence classes of problems they face. Molecular-scale descriptions begin by distinguishing parts of the processing system on the basis of neuroanatomy or biochemistry, or, in an AI setting, on the basis of hardware or encapsulated architectural modules. In AI, designers often deliberately simplify systems by ensuring that molecular-scale and molar-scale boundaries do not cross-cut one another. In general, however, this is only possible in contexts where problems that the system must solve are artificially restricted (that is, for systems functioning in artificial or 'toy' environments). In systems that co-evolve with complex environments, molar-scale equivalence classes tend to be highly heterogeneous from the molecular point of view while remaining stable objects for scientific generalization due to external environmental pressures that 'capture' different molecular processes within distinctive patterns. The logic here is the same as that which explains convergence in evolution by adaptation to niches. At the level of phylogeny, the relevant external pressures are ecological; in the case of people they are mainly social, and frequently institutional. Philosophers have discussed this under the rubric of 'the multiple realization of the mental' (Putnam 1975). In cognitive science it is reflected in Marr's (1982) highly influential methodology. According to this approach, cognitive architectures at the 'algorithmic' level link molar-scale 'computational' level descriptions with independently discovered molecular-scale 'implementation' level descriptions of hardware or brain anatomy.

In the case of the models of intertemporal individual choice that interest economists, the most accurate verdict at the present time would be that the methodological sophistication of Marr's framework is not yet reflected, though one of the pioneers of neuroeconomics, Paul Glimcher (2003) has specifically urged its adoption. The four families of existing models will now be described, with a view to substantiating this verdict. We will see that specifications of cognitive architectures are largely missing. In light of this, it is hardly surprising that no beginning at all has yet been made on the further step of embedding economic models of inconsistent individual choice in social settings, as called for by, among others, Ross (2005, 2008) and Castelfranchi (2006).

A first family of models emerges from the students and intellectual descendents of psychologist Richard Herrnstein, and has been dubbed 'picoeconomics' by its most systematic exponent, George Ainslie (1992, 2001). It denotes applications of game theory to model procrastination / impulsive consumption phenomena as sometime equilibrium outcomes of games played amongst sub-personal interests. These interests are not derived from either molar-scale descriptions or models of cognitive architecture. Instead, they are constructed directly as abstract manifestations of the pursuit of goals attributed at the personal scale and then hyperbolically discounted from points of choice. Thus, for example, a person trying to quit smoking has a short-range interest in having a cigarette and a long-range interest in not having one. These interests interact strategically. Thus the short-range interest might strengthen its prospects by allying with and promoting an interest in going to the bar, where a smoking lapse is more likely, while the longer-range interest might advance its cause by teaming up with an interest in going jogging. Hyperbolic discounting emerges to the extent that the short-range interest is not suppressed by defense of 'personal rules' that re-frame choices as between temporally extended *sequences* of alternatives. The successful quitter, on Ainslie's account, recognizes that a triumph of the short-run interest today predicts future such triumphs in similar future circumstances. This can lead her to notice that her effective choice is not between a cigarette now and one fewer lifetime cigarette, but between a series of indulgent choices and a series of prudent choices. The value of the latter may outweigh the value of the former from the current perspective *despite* hyperbolic discounting, because more valuable future prudent choices are summed by the re-framing, a perception that Ainslie calls 'reward bundling'. The bundler then chooses to pay a cost in the form of voluntary suffering from restraint, which would be pointless if relapse were sure. In economic terms, such behavior is a form of investment.

Note that on Ainslie's model, willpower is a matter of becoming aware of, and maintaining salience of, certain *information* – specifically, the information that present choices carry about future choices. In light of this, the picoeconomic model has the same structural character as Prelec & Bodner's (2003) *self-signaling* model. The variation introduced by Prelec & Bodner is that the information that triggers bundling is not so much about the choice situation as about the agent herself: the bundler's prudent choice informs her that she is the sort of agent who makes prudent choices and thus will likely make more such choices in the future, whereas impulsive choice signals the opposite sort of character.

Another framework that is logically akin to picoeconomics, though it does not use the label, has been introduced by the late behavioral economist Michael Bacharach (2006). This understands prudent choice as involving 'team reasoning', in which diachronically distinct sub-selves involved in games with one another re-frame their very agency so as to identify with one another. Formally, this changes the relevant games and so may change equilibrium outcomes. A present self that is in a Prisoner's Dilemma (PD) with future selves, and therefore defects by choosing impulsively, transforms the PD into an assurance game if it frames itself as part of a team with future selves and aims to maximize team utility. The assurance game model, unlike the PD, has Nash equilibria in which present selves cooperate with future selves by making prudent choices.

A cognitive scientist or AI researcher would be likely to ask for, or try to develop, a cognitive architecture that implemented Ainslie's and / or Bacharach's molecular-scale accounts. To date, however, such work has not been forthcoming. We may speculate that this reflects the intellectual origins of picoeconomics in the moderate behaviorism associated with Richard Herrnstein, who taught Ainslie and many of his leading interlocutors. There is no reason to suppose that Ainslie, Prelec or other currently active students of Herrnstein would be *opposed to* their models being set into specific cognitive architectures; the point is merely that this is not the kind of modeling that has an established place in the research tradition of which they are members.

The second family of models of impulsive choice and control, and the family that is most popular among economists, is based on β-δ discounting. These models also parse the agent into sub-agents, with the difference that, in any given choice problem, one sub-agent can be identified with the β discounter and another sub-agent with the δ discounter. This means that each agent is an exponential discounter as in equation (1) above, so standard economic techniques for identifying unique optimization equilibria can be applied. Examples of models that apply this approach are O'Donoghue & Rabin (2001), Bénabou & Tirole (2004), Benhabib & Bisin (2004), Berheim & Rangel (2004), and Fudenberg & Levine (2006). Recently some neuroeconomists and behavioral economists have exploited β-δ discounting to interpret neuroimaging data in support of a molecular-scale account of intertemporal preference reversal. McClure *et al* (2004, 2007) obtained fMRI evidence they interpret as suggesting that evolutionarily older orbitofrontal brain areas discount more steeply than later-evolving prefrontal areas. They then propose that hyperbolic discounting at the molar scale be understood as an aggregation of the tug of war between neurally localized β-discounting orbitofrontal and δ-discounting frontal sub-agents. This interpretation has lately been very widely embraced and promoted by behavioral economists.

The McClure *et al* hypothesis does not include, in Marr's terms, an algorithmic-level account. Its focus is entirely at the implementation level, where it suggests that firing rates in different groups of neurons respectively implement β and δ discounting. The hypothesized 'tug of war', which can presumably be influenced by cognitive structures and social influences, would need to be modeled at the level of cognitive architecture. Since these processes are black-boxed in the current model, it offers no immediate purchase on the motivations and conditions for reward bundling, as pointed out by Ross *et al* (2010).

The dual brain center interpretation of the McClure *et al* findings has encountered direct empirical difficulties reported by Glimcher *et al* (2007), which are set in a wider context by Ross *et al* (2010), who conclude that the weight of evidence is against the hypothesis. However, what is most relevant in the present context is the clear preference among behavioral economists for an explanation of the molar-scale behavioral pattern that appeals to the implementation level and black boxes the algorithmic level.

The third family of models conceptually shrinks the processes identified with the agent, treating parts of a person's brain as generating exogenous environmental impacts on the agent. Allowing for important variations in details, this modeling approach is shared by Loewenstein (1996, 1999), Read (2001, 2003), and Gul & Pesendorfer (2001). These models (of which only Gul and Pesendorfer's are fully explicit in economic terms) all explain personal-scale violations of thin economic rationality as resulting from 'visceral' temptations to immediately consume certain sorts of rewards, which the agent may or may not successfully resist. In these models, resisting temptation is expensive for agents (paid for in short-range suffering), but so is succumbing (paid for in lower longer-range utility). Thus the appearance of a temptation constitutes a negative shock along the agent's optimizing path. How agents respond to such shocks is a function of relative costs, which agents minimize subject to an exponential discount function. The resulting behavioral pattern, if graphed as though it were all just discounting behavior, yields a β-δ curve. This family of models implicitly denies any mediating role for cognitive architecture between the economic and molecular accounts, because the former is a purely abstract description of a relationship between price changes and behavior. It is not, in general, evident that Marr's picture of cognitive explanation applies to Gul & Pesendorfer's model. This would of course be consistent with their denial, in their methodological work, that economists and cognitive or brain scientists are interested in the same phenomena as one another.

Finally, a fourth style of modeling is urged by Glimcher (2009). He defines a new, distinctively neuroeconomic, concept he calls *subjective value* (SV), which resembles the tradition economist's idea of utility in respects *other than* those relevant to its social and welfare properties. The units of SV are action potentials per second, defined as the mean firing rates of specific populations of neurons. SVs are hypothesized as being always stochastically consistent with choice, even when expected utilities are not. Candidate axioms for specifying SV are proposed, though Glimcher is explicit that he expects these to be modified under empirical pressure. The hypothesis is rendered into a specifically neuroscientific one by Glimcher's suggestion that the weight of fMRI and other evidence indicates that the SV of an action or good is encoded by the mean activity in the medial prefrontal cortex and the ventral striatum. He speculates that the former encodes goods valuation and the latter encodes action valuation.

The suggestion that SV reliably predicts molar-scale choice gives Glimcher's hypothesis a reductionist character. Ross *et al* (2008) criticize this aspect of it. Molar-scale utility, they argue in deference to models of ecological rationality discussed above, may be partly calculated by the person in conjunction with external systems in the environment. Thus it may systematically diverge from value-in-the-brain. The picoeconomist grants that SV or something much like it may be a crucial *input* to personal choice, but also allows for other sorts of input, many of which are simply 'choice-governing tracks' laid out in the subject's cultural / social / market environment and which the brain may never explicitly evaluate. Of course, brains must always do *something* to produce behavior that implements choices; but this may not generally, let alone always, be direct neural computation of comparative reward values.

All of the modeling approaches to intertemporal discounting reviewed so far share the assumption that there is a single best functional form for use in modeling. This assumption is strongly called into question by recent experiment work reported in Andersen *et al* (2008, 2010). They estimate a so-called 'mixture model' that allows sufficiently rich data to empirically indicate proportions of intertemporal choices best captured by exponential, hyperbolic and β-δ functions. Using a battery of experiments with a representative sample of over 400 adult Danes, they acquired data of sufficient richness to permit structural estimation of the theoretical models, allowing for joint identification of utility functions and discount functions. They found much shallower discount rates than are typically reported by behavioral economists who do not control for attitudes to risk and the convexity of utility functions. This should not be regarded as an unexpected discovery, but rather as empirical confirmation of a methodological shortcoming in most behavioral economics. At least as strikingly, they found no evidence to support β-δ models, only modest evidence that any choices were best modeled by hyperbolic functions, and no evidence that such hyperbolic discounting as was supported made a substantial difference in economic magnitudes.

The survey of efforts to settle on models of intertemporal reward discounting for incorporation in larger economic models indicates general and representative failure to map a readily navigable frontier between cognitive science and economics. The final section of the chapter will offer reflections on these.

3. General discussion

Ross (2005) and Ross *et al* (2008) argue that the main factors that assist people in avoiding impulsive consumption – that is, in approximating traditional, intertemporally consistent economic agency – are social structures and the judgments and expectations of other people. This account of the sample phenomenon from Section 2 is broadly in accord with Smith's (2007) emphasis on ecological rationality. It is also congruent with Castelfranchi's (2006) remark that "models [of architectures of mind] should not only 'compete' but should also be compared to and integrated with the models of the other sciences more 'entitled' to model mind or society" (p. 357). Castelfranchi follows this with comments that pertain more specifically to economics:

> [The] AI approach to cognitive architectures – as not identical to psychological models – will for example … avoid the bad alternative currently emphasized in economic studies between (Olympic or bounded) formal, simple models of rationality and too empirically oriented, data-driven, non-formal and non-general models of human economic choices based on experiments (behavioral economics) (Ibid).

Part of the claim here is that cognitive science cannot be expected to integrate usefully with any economics that restricts its attention entirely to axiomatic theories and reduced-form models. It is not necessary to endorse currently popular, Luddite hostility to such models (as shown by, e.g., Moss 2006) in order to agree with this. As noted earlier, Gul & Pesendorfer (2008) make the same point from the other side. With respect to the other alternative mentioned by Castelfranchi, the material reviewed in this chapter demonstrates the limitations of standard behavioral economics: notwithstanding its devotion to socially mediated preferences, it cleaves too tightly to methodological individualism by trying to locate economic valuation and choice entirely 'within the agent's head'. The same point applies to the rival neuroeconomic approaches of McClure *et al* (2004, 2007) and Glimcher (2009).

So Castelfranchi's judgment is sound. However, he doesn't mention what has recently come to be the most common methodology in the parts of economics that closely border cognitive science and psychology with respect to target explanatory phenomena. This is to estimate and econometrically test models with varying, but typically non-linear, structures, and featuring statistical choice patterns as dependent variables. In this framework, the only natural path to integration with cognitive science would be identification by cognitive scientists of independent variables for inclusion on the right-hand sides of econometric models.

This immediately raises problems. Models of cognitive architectures are *processing* models. They are typically evaluated by qualitatively or quantitatively comparing their outputs to human or animal behavior. They seldom come attached with indicator variables, of the sort that could be embedded in econometric models. Furthermore, economic models that avoid the charge of implicit methodological individualism, for example the Andersen *et al* model discussed at the end of Section 2, generally do so by aggregating choices of many individuals. The Andersen *et al* criticism of non-standard discounting functions is mute on the question of whether the heterogeneity they observe is between individuals or within individuals; thus it is unclear what kind of cognitive model one might best compare with it.
In general, economists other than behavioral economists are more interested in aggregated than in individual behavior. Thus one might suppose that the most fruitful interanimations of economic and cognitive studies will involve multi-agent simulations. As pointed out in Section 1, however, to date such simulations of economic phenomena have involved only simple agents. Interesting dynamics in such models are social rather than cognitive.

It must thus be concluded that cognitive science has not yet become a significant supplier of variables or parameters to constrain economic models. Future, more productive interdisciplinary collaboration likely depends on progress in greater integration of cognitive models and multi-agent models of social interaction more generally (Sun 2006). However, we must sound a note of caution due to the fact that the trend among economists who are interested in ecological rationality is to model agents as driven more by institutional structures and less by inboard computations. The parallel rise of neuroeconomics has not incorporated interest in Marr's algorithmic level – precisely the ground on which peoples' strategic problems would be considered in tandem with biological restrictions on the computations they perform.

Perhaps in decades ahead a survey of the role of cognitive variables in economics will report a very different set of contributions and conclusions. There may be considerable space open for methodologically innovative researchers. They must begin, however, from recognition of the reasons that have kept the space open.

It is hoped that the present chapter points toward progress by helping to consolidate these reasons.

References

Ainslie, G. (1992). *Picoeconomics*. Cambridge: Cambridge University Press.

Ainslie, G. (2001). *Breakdown of Will*. Cambridge: Cambridge University Press.

Andersen, S., Harrison, G., Lau, M., & Rutström, E. (2008). Eliciting risk and time preferences. *Econometrica* 76: 583-619.

Andersen, S., Harrison, G., Lau, M., & Rutström, E. (2010). Discounting behavior: A reconsideration. Working Paper 2011-03, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, http://cear.gsu.edu/papers/index.html.

Anderson, P., Arrow, K., & Pines, D. (eds.) (1988). *The Economy as an Evolving Complex System.* Boston: Addison-Wesley.

Angner, E., & Loewenstein, G. (forthcoming). Behavioral economics. In U. Mäki (Ed.), *Handbook of the Philosophy of Science, Volume 13: Philosophy of Economics.* Amsterdam: Elsevier.

Ariely, D. (2008). *Predictably Irrational*. New York: Harper Collins.

Arrow, K., & Debreu, G. (1954). Existence of equilibrium for a competitive economy. *Econometrica, 22*, 265-290.

Arthur, W.B., Durlauf, S., & Lane, D. (eds.) (1997). *The Economy as an Evolving Complex System II*. Boston: Addison-Wesley.

Bacharach, M. (2006). *Beyond Individual Choice.* Princeton: Princeton University Press.

Becker, G. (1962). Irrational behavior and economic theory. *Journal of Political Economy*, 70: 1-13.

Becker, G., & Murphy, K. (1988). A theory of rational addiction. *Journal of Political Economy, 96*, 675-700.

Bénabou, R., & Tirole, J. (2004). Willpower and personal rules. *Journal of Political Economy, 112*, 848-886.

Benhabib, J., & Bisin, A. (2004). Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption-saving decisions. *Games and Economic Behavior, 52*, 460-492.

Bernheim, B. D., & Rangel, A. (2004). Addiction and cue-triggered decision processes. *American Economic Review, 94*, 1558–1590.

Bernheim, B.D., & Rangel, A. (2008). Choice-theoretic foundations for behavioral welfare economics. In A. Caplin and A. Schotter, eds., *The Foundations of Positive and Normative Economics: A Handbook*. Oxford: Oxford University Press, pp. 155-192.

Bicchieri, C. (2005). *The Grammar of Society.* Cambridge: Cambridge University Press.

Binmore, K. (2009). *Rational Decisions.* Princeton: Princeton University Press.

Blume, L., & Durlauf, S. (eds.) (2005). *The Economy as an Evolving Complex System III.* Oxford: Oxford University Press.

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

Castelfranchi, C. (2006). Cognitive architecture and contents for social structure and interactions. In R. Sun, ed., *Cognition and Multi-agent Interaction*, Cambridge: Cambridge University Press, pp. 355-390.

Clark, A. (1997). *Being There*. Cambridge, MA: MIT Press.

Conte, R., & Castelfranchi, C. (1995). *Cognitive and Social Action*. London: UCL Press.

Conte, R., & Paolucci, M. (2002). *Reputation in Artificial Societies*. Dordrecht: Springer.

Cressman, R. (2003). *Extensive Form Games and Evolutionary Dynamics.* Cambridge, MA: MIT Press.

Debreu, G. (1959). *Theory of Value.* New York: Wiley.

Debreu, G. (1974). Excess demand functions. *Journal of Mathematical Economics*, *1*, 15-23.

Fudenberg, D., & Levine, J. (2006). A dual-self model of impulse control. *American Economic Review, 96*, 1449-1476.

Gigerenzer, G., Todd, P., & the ABC Research Group (1999). *Simple Heuristics that Make Us Smart.* Oxford: Oxford University Press.

Glimcher, P. (2003). *Decisions, Uncertainty and the Brain.* Cambridge, MA: MIT Press.

Glimcher, P. (2009). Choice: Towards a standard back-pocket model. In P. Glimcher, C. Camerer, E. Fehr and R. Poldrack, eds., *Neuroeconomics: Decision Making and the Brain*. London: Elsevier, pp. 503-521.

Glimcher, P., Camerer, C., Fehr, E., & Poldrack, R., eds. (2009). *Neuroeconomics: Decision Making and the Brain*, London: Elsevier, pp. 323-329.

Glimcher, P., Kable, J., & Louie, K. (2007). Neuroeconomic studies of impulsivity: Now or just as soon as possible? *American Economic Review, 97*, 142-147.

Gode, D., & Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: market as a partial substitute for individual rationality. *Journal of Political Economy* 101(1): 119-137.

Grassé, P. (1959). La reconstruction du nid et les coordinations individuelles chez *Bellicositermes natalensis* et *Cubitermes* sp. La théorie de la stigmergie: Essai de'interprétation du comportement des termites constructeurs. *Insectes Sociaux, 6*, 41-83.

Gul, F., & Pesendorfer, W. (2001). Temptation and self control. *Econometrica, 69*, 1403-1436.

Gul, F., & Pesendorfer, W. (2008). The case for mindless economics. In A. Caplin and A. Schotter, eds., *The Foundations of Positive and Normative Economics: A Handbook*, Oxford: Oxford University Press, pp. 3-39.

Harrison, G. (2008). Neuroeconomics: A critical reconsideration. *Economics and Philosophy, 24*, 303-344.

Harsanyi, J. (1967). Games with incomplete information played by "Bayesian" players I-III. *Management Science* 14: 159-182.

Heyman, G. (2009). *Addiction: A Disorder of Choice*. Cambridge, MA: Harvard University Press.

Hutchins, E. (1995) *Cognition in the Wild.* Cambridge, MA: MIT Press.

Jevons, W. S. (1871). *The Theory of Political Economy.* London: Macmillan.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heursitics and Biases.* Cambridge: Cambridge University Press.

Laibson, D. (1994). *Hyperbolic Discounting and Consumption*. Ph.D. Thesis, MIT.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics, 6*, 443-479.

Lichtenstein, S., & Slovic, P. (2006). *The Construction of Preference.* Cambridge University Press.

Lipsey, R., & Lancaster, K. (1956). The general theory of second best. *Review of Economic Studies*, *24*, 11-32.

Loewenstein, G. (1996). Out of control: visceral influences on behavior. *Organizational Behavior and Human Decision Processes, 65*, 272-292.

Loewenstein, G. (1999). A visceral account of addiction. In J. Elster and O.-J. Skog, eds., *Getting Hooked: Rationality and Addiction*. Cambridge, MA: Cambridge University Press, pp. 235-264.

Mantel, R. (1974). On the characterization of aggregate excess demand. *Journal of Economic Theory*, *7*, 348-353.

Mantel, R. (1976). Homothetic preferences and community excess demand functions. *Journal of Economic Theory*, *12*, 197-201.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.

McClure, S., Ericson, K., Laibson, D., Loewenstein, G., & Cohen, J. (2007). Time discounting for primary rewards. *The Journal of Neuroscience, 27*, 5796-5804.

McClure, S., Laibson, D., Loewenstein, G., & Cohen, J. (2004). The grasshopper and the ant: Separate neural systems value immediate and delayed monetary rewards. *Science, 306*, 503-507.

Menger, C. (1871 / 1981). *Principles of Economics*. Trans. J. Dingwall & B. Hoselitz. New York: New York University Press.

Moss, S. (2006). Cognitive science and good social science. In R. Sun, ed., *Cognition and Multi-Agent Interaction*. Cambridge: Cambridge University Press, pp. 393-400.

Musau, A. (2009). Modeling alternatives to exponential discounting. MPRA Paper No. 16416. Online at http://mpra.ub.uni-muenchen.de/16416/

O'Donoghue, T., and Rabin, M. (2001). Choice and procrastination. *Quarterly Journal of Economics, 116*, 121-160.

Pareto, V. (1909 / 1971). *Manual of Political Economy.* New York: Augustus Kelley.

Phelps, E. & Pollack, R. (1968). On second-best national saving and game-equilibrium growth. *Review of Economic Studies, 35*, 185-199.

Prelec, D., & Bodner, R. (2003). Self-signaling and self-control. In G. Loewenstein, D. Read & R. Baumeister, eds., *Time and Decision*. New York: Russell Sage Foundation, pp. 277-298.

Putnam, H. (1975). *Mind, Language and Reality*. Cambridge: Cambridge University Press.

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of Risk and Uncertainty, 23*, 5-32.

Read, D. (2003). Subadditive intertemporal choice. In G. Loewenstein, D. Read and R. Baumeister, eds., *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*. New York: Russell Sage Foundation, pp. 301-322.

Robbins, L. (1935). *An Essay on the Nature and Significance of Economic Science,* 2nd edition. London: Macmillan.

Ross, D. (2005). *Economic Theory and Cognitive Science: Microexplanation.* Cambridge, MA: MIT Press.

Ross, D. (2008). Economics, cognitive science and social cognition. *Cognitive Systems Research, 9*, 125-135, 2008.

Ross, D. (2010). Economic models of procrastination. In C. Andreou and M. White, eds., *The Thief of Time.* Oxford: Oxford University Press, pp. 28-50.

Ross, D., Ainslie, G., & Hofmeyr, A. (2010). Self-control, discounting and reward: Why picoeconomics is economics. Working paper, Center for the Economic Analysis of Risk, Georgia State University. Online at http://www.cear.gsu.edu/files/Ross.pdf

Ross, D., Sharp, C., Vuchinich, R., & Spurrett, D. (2008). *Midbrain Mutiny*. Cambridge, MA: MIT Press.

Samuelson, P. (1938). A note on the pure theory of consumer's behavior. *Economica, 5*, 61-72.

Samuelson, P. (1947). *Foundations of Economic Analysis.* Enlarged edition (1983). Cambridge, MA: Harvard University Press.

Savage, L. (1954). *The Foundations of Statistics.* New York: Wiley.

Simon, H. (1969). *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Simon, H. (1972). Theories of bounded rationality. In C.B. McGuire and R. Radner eds., *Decision and Organization*. Amsterdam: North-Holland, pp. 161-76.

Smith, V. (2007). *Rationality in Economics*. Cambridge: Cambridge University Press.

Sonnenschein, H. (1972). Market excess demand functions. *Econometrica*, *40*, 549-563.

Sonnenschein, H. (1973). Do Walras identity and continuity characterize the class of excess demand functions? *Journal of Economic Theory*, *6*, 345-354.

Sun, R. (2006). Prolegomena to integrating cognitive modeling and social simulation. In R. Sun, ed., *Cognition and Multi-agent Interaction*, Cambridge: Cambridge University Press, pp. 3-26.

Sunder, S. (2003). Market as artifact: aggregate efficiency from zero intelligence traders. In M. Augier and J. March, eds., *Models of a Man: Essays in Memory of Herbert A. Simon*. Cambridge, MA: MIT Press, pp. 501-519.

von Mises, L. (1949). *Human Action: A Treatise on Economics*. New Haven: Yale University Press.

von Neumann, J., & Morgenstern, O. (1944). *The Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Weibull, J. (1995). *Evolutionary Game Theory.* Cambridge, MA: MIT Press.

Zak, P. (2008). The brains behind economics. *Journal of Economic Methodology, 15*, 301-302.