THE ECONOMIC AGENT: NOT HUMAN, BUT IMPORTANT

Don Ross

1 INTRODUCTION

Critics of mainstream economics typically rest important weight on the differences between people and the 'agents' that populate economic theory and economic models. Hollis and Nell [1975] is both representative of and ancestral to many more recent variations on the theme. Lately, the upgraded status of behavioral economics (BE) within the discipline's mainstream has encouraged a number of writers to use revolutionary rhetoric in promotion of a 'paradigm shift' that includes the rejection of 'rational economic man' [Ormerod, 1994; Heilbroner and Milburg, 1995; Fullbrook, 2003]. The current leading developers of BE are generally more circumspect, claiming that their approach complements standard theory rather than promising to supplant it [Camerer and Loewenstein, 2004; Angner and Loewenstein, this volume]. However, they generally join the more florid critics in supposing that microeconomics is bound to improve its empirical relevance to the extent that it substitutes the study of people for that of abstract economic agents. Another body of thought that promotes this view stems from Sen's [1977] attack on standard economic agents as 'rational fools', amplified in Davis's [2003] argument that since economic agents lack some *essential* properties of human individuals, economic theory requires fundamental reform if it is to make progress in explaining human behavior.

That economic agents and people have different properties should strike no one as surprising. Whereas people are pre-theoretical entities found in the world, economic agency is a theoretical construction elaborated as part of the development of a family of models. In philosophical terms, we might therefore describe the view that economists should forget about economic agency and directly study people instead as an expression of *normative phenomenalism*. This would be the thesis that the proper objects of scientific attention are manifest phenomena, which should be described directly rather than by way of intermediate theoretical kinds. This is not a view we find on display elsewhere in the philosophy of science. Though strong empiricists, such as Bas van Fraassen [1980; 2002], deny that we are entitled to ascribe model-independent reality to the unobservable objects of reference used in scientific theories, I have never heard anyone insist that physicists ought to stop modeling fields and manifolds and go back to generalizing directly about

Handbook of the Philosophy of Science. Volume 13: Philosophy of Economics.
Volume editor: Uskali Mäki. General editors: Dov M. Gabbay, Paul Thagard and John Woods.

rocks and tables. Elsewhere in this volume, I extensively discuss the reasons why economics has attracted a level of anti-theoretical *hostility* not encountered by other sciences (aside from evolutionary biology). I suggest that this discussion is useful background for explaining the eagerness with which revolutions in economics are promoted on grounds we don't encounter elsewhere in science. In the present essay, I will assume that normative phenomenalism, especially as applied arbitrarily to and only to economics, is not rationally motivated. This assumption does not foreclose the possibility that either current critique or the future course of economic science could reveal the idea of economic agency, either in general or in some common particular form, to be unhelpful. Use of any given agency concept in science *is* subject to requests for justification; but the mere fact that economic agency is abstractly constructed establishes *no* prima facie case against such justification.

It might be objected that normative phenomenalism is a fair standard for application to economics in particular because economic theory, unlike physical theory, generalizes only or mainly over observable types. In this connection, Mäki [1986; 1992] points out that Friedman's [1953] famous methodology isn't in fact the standard sort of instrumentalism it's typically taken to be because, unlike philosophical instrumentalists about the unobservable entities of physics, Friedman assumes the objects of economics to be manifest: consumers, firms, prices, etc. He then doubts that economic theory truly describes these objects, useful though it is for predicting their trajectories. He does not doubt, as the instrumentalist does of bosons, that the basic objects of economics exist. Though I agree with Mäki about what Friedman thought on this question, I do not think that Friedman's opinion here is correct. Because the *words* used by economists, unlike 'boson', are derived from everyday vocabulary, it is easy to forget that in their theoretical context they denote abstractions. Despite slightly quaint philosophical jargon, Stigum [1990] offers nice examples of the point: "We have knowledge by acquaintance of the salary we received last year, but we have knowledge by description only of what our income was, i.e., of the maximum amount of money we could have spent last year and been as wealthy at the end of the year as we had been at the beginning of the year ... We have knowledge by acquaintance of the price of our house, but only knowledge by description of its current market value" (p. 550). So it is with agency in economic theory: we gaze upon and shake hands with people, but not with economic agents. But, in the absence of an argument for normative phenomenalism, this fact by itself no more implies that economists should stop theorizing about agents, or equate them with people, then similar logic would rightly advise them to stop theorizing about incomes or to equate incomes with salaries. Again, this is not to deny the validity of requests for justification on grounds internal to the goals of economics (as opposed to external philosophical grounds).

The structure of the chapter is as follows. I will first sketch the standard concept of the economic agent as featured in contemporary microeconomics. I will then show why the practice of economists does not equate this agent to a person, and why economists' longstanding interests in 'individualism' and 'microfoundations'

should not be interpreted as suggesting otherwise. This will show how, in detail, economists should respond to criticisms reflecting normative phenomenalism. In section 5 I will indicate why and how (some) behavioral economists propose to modify agency in light of studies of people, in cases where normative phenomenalism is *not* assumed. The core of this argument involves contesting the view held by increasingly many behavioral economists that their program collapses into the ambition of the new 'neuroeconomics' to identify and explain the processes by which brains comparatively value actual and prospective rewards. I will maintain that what I will call 'neurocellular economics' (as found in work by [Glimcher, 2003; Caplin and Dean, 2008]) is importantly different in its implicit attitude to standard economic agency from a more reductionist version of neuroeconomics that has lately been stapled to BE in would-be service of a paradigm shift [Camerer *et al.*, 2005]. Having explained why modular neuroeconomics preserves rather than challenges the standard concept of economic agency, I will defend the continued use of that concept against calls for its replacement by objects and processes identified through psychological and neuroscientific observation.

2 ECONOMIC AGENCY

There is a clear historical path by which the standard concept of the economic agent was developed [Mandler, 1999]. This agent first appeared in the work of the early neoclassical theorists (Jevons and Walras) as a maximizer of ceiling-less hedonic utility laboring under a finite budget, subject to diminishing marginal returns from consumption within classes of commodities he deemed to be close substitutes. I deliberately use the pronoun 'he', because at the point of his historical arrival the economic agent was both normatively male in his status as a social atom and (more importantly for present purposes) human, in that he relied on 'creature sensations' to both form his close-substitute classes and to rank them with respect to the utility they delivered. His agency revolved around his efforts, given his limited means, to create the most appealing inner environment he could, as determined by his own introspective judgment.

Although the early neoclassical agent was human, he was already not a *whole* person. In sympathy with Mill's refusal to follow Bentham in regarding all sources of satisfaction — pushpin and poetry, a foot message and an end to poverty — as lying on a single commensurable scale, Jevons [1871] took the economic agent to be the *aspect* of the person concerned with the consumption of 'lower' wants. We can fully understand what was 'lower' about these wants only by going slightly outside the frame of Jevons's text and importing some knowledge of the Victorian world-view. To some extent the lowliness of economic wants lay in their materiality. But Victorian idealism was closely bound up with the morality of social obligation: material goods were 'low' in part because, unlike 'spiritual' goods, their consumption as sources of utility was private; 'higher' wants were higher in part because attending to them expressed commitment to public civilization. Given the importance of atomism as a property of the economic agent for which

he is widely rebuked by his critics (including contemporary ones), this point merits its emphasis. The Victorians were pointedly and self-consciously divided amongst themselves as regards metaphysical atomism versus holism, with the scientifically minded such as Jevons inclining to the former and most philosophers defending the latter. But neither Jevons nor Walras were *moral* atomists; both rejected the idea that a person should give his highest priority to what they regarded as his economic interests.

Some readers might have jumped to the conclusion that I am calling the early neoclassical agent 'human' because he had 'feelings'. It has long been fashionable to contrast the 'cold calculator' featured in economic theory with the warm, sentimental and impulsive beings celebrated by all romantics and by most Western humanists. Though this is important for understanding sources of non-rational antipathy to economics, it seems to me ethnocentric to view emotional parochialism and impulsiveness as the core properties of the human; Western romantic humanism is a peculiar, not a globally typical, idealization of human nature. Thus I would question the long-run philosophical importance of contrasting the early agent's passions with the later agent's lack of them. Instead, I suggest, what made Jevons's economic agent human by contrast with his contemporary successor was the former's grounding in consumption within the boundaries of his *body*. The early neoclassical agent was an aspect of the human *animal*. Thus there was an implicit one-to-one mapping between these agents and human organisms, which all applications took for granted.

As recognized by many writers, and reviewed succinctly by Bruni [2005] and Bruni and Sugden [2007], Jevons's introspective agent was on the way out before the twentieth century began; Pareto, in particular, worked to reduce his defining properties to a mere disposition to consume in accordance with representation by indifference curves. Following on this lead using more powerful mathematical resources, the introspective agent was killed stone dead in the ordinalist revolution of the 1930s and 40s led by Hicks, Allen and Samuelson [Mandler, 1999]. As related by Ross [2005], however, what never disappeared from most economists' (or other people's) informal conception of the economic agent was the idea that he was still (as it were) 'ontologically grounded' in the human organism. By this I mean only that the one-to-one mapping between agents and organisms presumed by Jevons and Walras (henceforth, '$A \Leftrightarrow O$') remained the basic reference point for understanding the place of agents in the empirical interpretation of economic theory, even as the agent's human properties were steadily stripped away. There were motivations for this conservatism, as we will see; it wasn't merely a case of conceptual inertia. But, I will argue, we can make more consistent sense of the character of most economics since Samuelson by dropping the attribution to its foundations of the assumption of $A \Leftrightarrow O$.

The ordinalist revolution did not so much modify the concept of the economic agent as, to begin with, attempt to eliminate him. In the canonical ordinalist texts, Samuelson [1938; 1947] set out to derive the existence of sets of preferences mappable onto the real numbers by monotonic, complete, acyclical, and convex

functions from observable schedules of aggregate demand. He would have preferred not to call these 'utility' functions, but the lure of semantic continuity turned out to be a more powerful force than his preference, and he quickly surrendered the point to convention.[1] As the label 'revealed preference theory' was intended to suggest, his utility functions were intended as descriptions of actual and hypothetical behavior, not inner evaluations of experienced relative states of satisfaction. 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, imagining it to have been the main, let alone the sole, motivation ignores the fact that Samuelson completed a process that had been underway for decades in economics, and which thus reflected a special dynamic internal to the discipline. This was the felt pressure to make economics a *social* science independent of any foundations in individual psychology. Cold war neuroses demanding adherence to 'methodological individualism' did much to obscure the point in retrospect. But as good Keynesians, Hicks and Samuelson were, in a very important sense to which we will return later, *uninterested* in individual agents, a concept of which they merely inherited from an earlier neoclassical theory they profoundly transformed. If we let Samuelson's [1947] mathematics speak for itself, as he largely though inconsistently does himself in *Foundations*, then among the short and general things we might say about the role of the agent in revealed preference theory the most accurate is that there isn't one. There is observable aggregate demand, and if this has certain testable properties then the existence of continuous preference fields is implied. What stabilizes these fields might or might not be properties of individual psychologies; the revealed preference theorist disavows professional interest in this question, a point on which Samuelson is explicit.

All this makes it easy to imagine that, and how, the agent might have disappeared altogether from economic theory had the discipline technically matured in a slightly different context. Indeed, someone might well argue that the agent *did* substantially disappear despite the fact that the *word* 'agent' soon made a comeback in the literature following on Samuelson. There are three possible interpretations to be distinguished here. By 'interpretations' I refer not to claims about what historical economists actually intended, but to attributions that might be offered by philosophical reconstructions that apply retrospective principles of charity in full knowledge of contemporary economics. The possible interpretations are:

1. The role of the agent was eliminated from microeconomic theory after World War Two.

2. Postwar microeconomic theory retained a concept of the agent, but with substantial modifications that imply abandonment of the commitment to

[1] Many economists, however, now refer to 'objective functions' rather than 'utility functions'. I hope that this becomes standard usage, but fear that the influence of behavioral economics will get in the way.

$A \Leftrightarrow O$ (whether or not many economists, who are not in the philosophy business, noticed this).

3. The absence of agents in Samuelson's version of revealed preference theory was an idiosyncratic wobble in the evolution of microeconomic theory; the reappearance of the word 'agent' in subsequent canonical texts indicates stronger continuity with early neoclassicism than Samuelson suggested, in particular, continued ontological orientation around $A \Leftrightarrow O$.

Contemporary paradigm-shifters based in BE, along with Sen and his followers, adhere to interpretation (3) and then, in rejecting social atomism, take themselves to be calling for the overthrow of a historically unified neoclassical tradition. (Thus they often refer to the contemporary mainstream as 'Walrasian'.) I will defend interpretation (2).

Let us now hoist the target of the conflicting interpretations onto the table. Again, there can be no dispute that Samuelson's avoidance of the *word* 'agent' failed to stick as a practice: the subtitle of Rubinstein's [2006] elegant formulation of the core elements of microeconomic theory, which deserves to be regarded as authoritative on matters of current convention, is "*The Economic Agent*". I will summarize the part of Rubinstein's formulation that might plausibly be taken to be *definitive* of economic agency. This is the part that can be stated independently of any assumptions about representations or computations taken to be aspects of agents' psychologies; were such assumptions to be incorporated into the definition of agency then the question distinguishing the defenders of interpretations (2) and (3) above would necessarily be begged in favor of the latter. Note that the judgment about what to regard as 'definitive' that I will express below is mine, not Rubinstein's. Note also that Rubinstein's formulation reflects the consolidation of postwar consumer theory provided by Debreu [1959], rather than the less exact version found in Samuelson [1947]; this is a point that will be important in the later discussion.

The agent is a reference point for ascription of a utility function. Utility functions are constructed from preference functions or represent preference relations. A preference function or relation generalizes a series of answers to a series of evaluative questions about elements $x, y, \ldots, n$ of a set X, with one answer per question of the form 'x is preferred to y' ($x \succ y$), 'y is preferred to x' ($y \succ x$), or 'x and y are interchangeable in preference ranking' (I). Rubinstein shows that two forms of generalization are equivalent:

1. Preferences on a set X are a function f that assigns to any pair (x, y) of distinct elements in X exactly one of $x \succ y$, $y \succ x$, or I, restricted by two properties: (i) *no order effect*: $f(x, y) = f(y, x)$; and (ii) *transitivity*: if $f(x, y) = x \succ y$ and $f(y, z) = y \succ z$ then $f(x, z) = x \succ z$ and if $f(x, y) = I$ and $f(y, z) = I$ then $f(x, z) = I$.

2. A preference on a set X is a binary relation $\succ$ on X satisfying (i) *completeness*: for any x, $y \in X$, $x \succ y$or $y \succ x$; and (ii) *transitivity*: for any

$x, y, z \in X$ if $x \succ y$ and $y \succ z$ then $x \succ z$.

A utility function is a representation of a preference relation according to: $U : X \to \Re$ represents $\succ$ if for all $x, y \in X, x \succ y$ if and only if $U(x) \geq U(y)$.

If the foregoing is taken to restrict the conception of an agent then it follows that an agent's preferences are not lexicographic [Debreu, 1960]. This also follows from conceiving of preference relations as continuous. From Debreu [1954; 1960], any set of continuous preferences is represented by a continuous utility function.

The agent distributes her investments in alternative feasible states of the world in accordance with the weak axiom of revealed preference. I use a formulation of my own here instead of Rubinstein's: for two complete states of the world $x, y : x \neq y$, if the agent pays opportunity cost $c + y$ in exchange for x, then the agent will never pay opportunity cost $c + x$ in exchange for y. This implies that the agent's behavior will be consistent with the hypothesis that she maximizes a utility function according to which $U(x) \geq U(y)$.

When agents are located in markets where they encounter consumption problems, more is generally assumed of them. In particular, it is supposed that when they are faced with alternative investments in quantitatively measurable combinations of elements (bundles) from their utility functions, their preferences satisfy monotonicity (for any element $x \in X, x + \varepsilon \succ x$), continuity, and convexity (consumption behavior is consistent with representation by neoclassical indifference curves). Stronger assumptions, particularly that utility functions are differentiable, are typically added if we are concerned to show that a particular model of a consumer's optimization of consumption given a budget is explained by reference to her preferences. Note that economists are almost never moved by this concern except when they are engaged in explicit justification of abstract theory — that is to say, when they're not actually doing economics.

In light of the foregoing, our prior question about the ontological presumptions around agency in postwar economic theory comes down to this: what import should be attached to saying that a reference point for ascription of a utility function, as just defined, is an 'agent'? 'Reference point' here just means an element of some index constructed for a particular analytic exercise; so all the weight lies on concept of the utility function. It should be evident that what I identified earlier as the 'human' properties of Jevons's agent make no appearance in the definition. Nor, at least until the rise of BE, did they play any explicit role in interpretations of the formalism in applications. Now, there is no room for serious doubt that in the Western intellectual tradition the prototypical agent is the goal-pursuing aspect of a single person over the course of her biography from the dawn to the demise of her mature competence in practical reasoning [Ross, 2002]. The idea has a relatively clear and constant conceptual core from the work of Aristotle through Kant. From this perspective it should seem puzzling that Samuelson's avoidance of reference to agents didn't continue to be respected: reference points for ascriptions of utility functions don't seem particularly to resemble philosophers' agents. Why then is it standard practice we find Rubinstein reflecting in using 'the economic agent' as his subtitle in 2006?

In aiming to be empirical scientists, rather than members of the community of mathematicians who study constrained optimization,[2] economists necessarily suppose that their theory gives a general description of some class(es) of empirical phenomena. At the most crude level of description, there seem to be two alternatives here: the theory can be about people, or it can be about emergent systems of production, consumption and exchange, in a context of agnosticism about who or what the *ultimate* units of these activities are (*if* there need to *be* ultimate such units at all; see [Ladyman and Ross, 2007] for reasons to doubt this). Once the issue is put this way, it might be supposed that the answer to the question at the end of the previous paragraph is obvious: utility functions must be proxies for individual flesh-and-blood consumers lest we implicitly endorse mysterious 'group minds' that don't decompose into individual minds; methodological individualism follows from metaphysical atomism. If utility functions map one-to-one onto people for philosophical reasons, then in light of the same philosophical tradition according to which $A \Leftrightarrow O$, a theory of the utility function is a theory of the agent.

However, economists are usually reluctant to accept important professional doctrines simply on philosophical grounds, as they should be. One consequence of the public prominence of the Chicago School has been to greatly exaggerate the perceived commitment to methodological individualism in workaday economics. Agnosticism about microfoundations need not imply — as it certainly didn't for Keynes or Samuelson — endorsement of a transcendent Hegelian spirit which, in addition to thinking about itself and moving history along, also produces, consumes and trades. The respectable scientists who work today in complex systems theory (who are respectable as scientists regardless of whether one shares their confidence in their approach) believe in emergent processes and entities, behavior of which cannot be derived from behavior of their constituents *in vitro*, but generally do not believe that feedback-regulated dynamical systems are manifestations of Spirit. Of course, complex systems theory did not yet exist in the 1950s. But this didn't deter Samuelson from haughty indifference about the atomic material contents of the economist's structural black boxes. (For example, at one point in the *Foundations* [p. 87] he effectively implies that the firm in production theory is not a 'company' in the everyday sense, since the latter but not the former may make profits; but, he says, studying institutional contexts that allow companies to gather rents is not the economist's business. This would imply that it is also not in the economist's brief to say why people form companies in the first place.) The real liberator of economists from the ball-and-chain of microfoundations was Keynes, who enjoyed emphasizing that the concerns of the philosophers in whose company he had been intellectually trained were of no practical import in the dangerous concrete world where policy was called upon to keep revolution at bay. Keynes made economics both theoretically autonomous and professionally thrilling, and these two attractive aspects of the profession as it set about reorganizing the postwar order were closely related to one another. The conquering macroeconomists

[2]Rosenberg [1992] argues that that is in fact what economists are, whether they mean to be or not. I disagree.

of the Bretton Woods era were neither metaphysical atomists nor metaphysical holists; they were practical structuralists who left metaphysics to others.

I have already alluded, in my reference to 'Cold War neuroses', to one reason this golden moment didn't last. Opposing Stalinism obviously didn't rationally *require* that anyone swear fealty to methodological individualism; but war is no friend to subtlety (nor, as emphasized by Mirowski [2002], were the military funding sources that fueled the expansion of postwar science, including economics[3]). It cannot be rigorously demonstrated, but nevertheless seems very likely, that extra-theoretical political factors in the postwar democracies constituted the most decisive influence on economists' return to the *rhetoric* of social atomism. Because such rhetoric was also widely associated — by the loosest, Humean, kind of relation — with defense of markets against 'collectivists', and because economists are indeed appreciators of markets, Chicago School celebrities readily promoted the idea that economic theory has both descriptive and normative individualism built into its core.

Though I contend that this was indeed more a matter of rhetoric than logic, it would be seriously mistaken to suppose that the only reason economic theory didn't continue down Samuelson's agent-free path is the purely external, sociological one that its popular image was captured by cold warriors. In the first place, as I argue elsewhere in this volume, the completeness of the capture is often exaggerated. In the second place, economists were not unaware that most of their applied work continued to focus on aggregate magnitudes and relations. Economists had *reasons*, grounded in microeconomics rather than metaphysics, for thinking that agency couldn't be excised from their theoretical foundations. I will concentrate on two.

First, the invention of game theory (GT) by von Neumann and Morgenstern in 1944 allowed economists to model the interactions of idiosyncratically varying utility functions rendered interdependent by contingent distributions of scarcity. Nothing in the mathematics stipulates that these must be interpreted as the utility functions of *people*; indeed, in the most useful contemporary *economic* (as opposed to psychological) applications of GT, they represent objectives of firms rather than of humans [Ghemawat, 1998; Klemperer, 2004; Milgrom, 2004]. However, GT required the enrichment of utility theory that von Neumann and Morgenstern (and then Savage) 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 heuristics drawn from folk psychology, and thus the non-mathematical version of the vocabulary of game theory is full of psychological notions: beliefs, conjectures, aversion, attraction. Furthermore, and more substantively, GT made it possible for economists to use the core elements

[3]In echoing Mirowski here, I intend to cast no aspersion on Cold War era economists. Fully morally reasonable scientists who are passionate about their subject matter should be *expected* to make non-vicious political compromises when unprecedented resources for their work flood around them. Had economists not been influenced by the interests of the postwar military there would have been something seriously wrong with the extent of their dedication *as scientists*. Of course, some will dispute my suggestion that most of the relevant compromises were non-vicious. That discussion must be left to another occasion.

of their conceptual toolkit (constrained optimization and opportunity cost) to systematically study individual choices in strategic contexts and so, like good opportunistic scientists, they duly embarked on such study. If we are to base our views of disciplinary boundaries on what scientists actually do instead of on philosophical doctrines about how the world is objectively carved, then we must agree that the early game theorists thereby widened the scope of economics, regardless of whether a revealed-preference purist would approve.[4] Finally, GT seemed to demand progressive deepening of links between economics and psychology as it technically evolved over the past 35 years. 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 interpret 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. As pointed out in criticism by Binmore [1990], the resulting 'refinement program' draws game theorists not just into psychology but deep into *philosophy*, since it requires them to study their own 'intuitions' about which chains of argument must be pursued if an agent is to count as 'rational'. In this context the idea of agency looks *fundamental* to microeconomics.

Second, the formal completion of general equilibrium theory by Arrow and Debreu [1954] required the concept of an 'economy' to be strictly regimented, and this in turn demanded imposition of strong general constraints on 'participants' in such economies [Debreu, 1959]. In particular, it was necessary to assume that the participants could rank all possible states of the world with respect to value, and that they never change their minds about these rankings. Again, nothing required that 'participants' be interpreted as coextensive with people. As argued at length in Ross [2005], if agents in general equilibrium are identified with utility functions, then the fact that changes in utility functions imply changes in agent identity is an excellent reason *not* to identify such agents with people. However, an important part of the intended point of general equilibrium theory, all the way back to Walras, has been to serve as a framework for thinking about the consequences of changes in exogenous variables, especially policy variables, for welfare. Regardless of whether descriptive individualism is persuasive as social *metaphysics* — the reader will have gathered that I think it is not — there remain the best of reasons for endorsing *normative* individualism: improvements and declines in the feelings of particular people about their well being is what most people, as a matter of fact, mainly care about, so for an economist to regard anything *else* as the appropriate topic of welfare analysis is to implicitly impose the economist's parochial value

[4]Thanks to Erik Angner for stressing this point to me.

scheme on society. Policy makers should ignore the advice of such economists.[5] Thus if the loci of preference fields in general equilibrium theory are not at least idealizations of *people*, then it is not evident why efficiency, the touchstone of general equilibrium analysis, should be important enough to *warrant* touchstone status.

Theoretical developments in the 1970s added economic substance to this philosophical concern. The 'excess demand' literature of that period, centering around the Sonnenschein-Mantel-Debreu theorem [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 directly, 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, Sonnenschein-Mantel-Debreu 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 (and aggregate-scale, sociological rather than psychological) 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 properties of individuals could be safely ignored.

Both the theory of individual choice under uncertainty and welfare theory are *extensions* of core microeconomic theory. Therefore, the fact that both embroil economists in issues about agency is not a slam-dunk argument for interpreting that core using the standard semantic label chosen by Rubinstein. However, here it is important to remember that if the pressure to regard economics as being about agents isn't decisive, the basis for resistance to such an interpretation isn't very powerful either. As observed above, in denying that macroeconomics had necessarily to be derived from microeconomics, Keynesians expressed commitment to pragmatism, not philosophical holism: they left microeconomics behind (Keynes) or blithely cast aside its early neoclassical commitments (Hicks and Samuelson) because they thought that rigid fealty to Jevons and Walras stood in the way of exercising available capacities to control policy-relevant economic relationships and magnitudes. Therefore, if we come around to the view that psychologistic GT is relevant to policy, as all behavioral economists believe, then the same attitude that led Samuelson to drop agents from his foundations should inspire us to put it back. Furthermore, if psychologistic GT is relevant to policy because of variations in individuals' utility functions and attitudes to risk, then it seems our idea of welfare

[5]I do not mean here to just *dismiss* views of those, such as Sen [1999], who think that people's subjective preferences are often unreliable guides to their well being (though I am suspicious of such views). The intended targets of this remark are critics, such as radical environmentalists, who believe that something other than the welfare of particular human beings is the most appropriate basic standard of valuation. In my opinion this requires an unsustainable level of moral arrogance, and is especially unpalatable when promoted by materially comfortable people in a world suffering from significant levels of true poverty.

is implied to be richer than merely the vague utilitarian commitment to maximize community indifference curves that characterizes most economics applied at the scales of national and international policy.

I think that these considerations do defeat interpretation (1) of the place of agency in postwar economic theory. Economics is motivated by a broader set of empirical observations than merely noticing that ecologies of self-maintaining entities collectively demand more consumption goods than the world can provide; it is equally fundamental to the discipline as we now find it that these entities have available to them and use importantly different strategy sets and strategies for coping with specific aspects of their scarcity problems. Once we have got as far as talking about 'entities with varying utility functions and strategy sets' then it would simply be conceptually obtuse to deny that our focus is on agents. Indeed, we should arrive at this conclusion with some relief. It spares us the need to try to make general sense of preference or consumption while not being able to say that there is any kind of thing that is, in general, a possible locus for having preferences and consuming. Let me be careful in framing the significance of this point. I don't wish to make philosophy seem too important here, and I don't believe that we can aspire to close the whole conceptual system by reducing basic economic concepts to some extra-economic bedrock. Instead, preference, consumption and agency, operationalized together as a triad, plausibly constitute a collective conceptual primitive for economics, and as long as this doesn't leave economics stranded apart from other sciences this should be regarded as foundations enough. My point here is just that leaving agency in the picture doesn't seriously compromise foundational elegance *given that* preference and consumption are already admitted. Therefore, *declining* to identify utility functions with agents would give *more* weight to philosophy — refusing to 'say what comes naturally', just out of philosophical scruples — than doing so.

However, giving up the radical ambition to eliminate agency from economic theory need not carry us, with Sen and the behavioral economists, all the way to interpretation (3). I will argue over the course of the remaining sections of the chapter that although economics is about agents, it is not best regarded as staked to $A \Leftrightarrow O$.

Before I launch into this, let me deflect a potential charge that I have announced battle with a straw opponent. It might be objected that the paradigm shifters have no need to accept a generalization as strong as $A \Leftrightarrow O$, and, indeed, *do not* insist on it. They will agree that many applications of economics treat firms, households, unions and even countries as agents. Furthermore, they will note — indeed, will emphasize — that models inspired by neuroeconomics focus on sub-personal agents [Montague *et al.*, 2006, p. 438]. This idea of representing people as communities of agents — synchronic, diachronic or both at once — goes back to the very dawn of BE [Strotz, 1956], and so has some claim to being regarded as among its basic points of departure from neoclassicism.

These points are duly acknowledged. I do not claim that any economists of note maintain $A \Leftrightarrow O$ as an analytic or metaphysical necessity. They are thus

open to extending the concept of agency to apply it to entities other than whole individual people, and they do regularly so extend it. However, my key point is precisely that *behavioral* economists must regard these as *extensions*. They join classical economists and early neoclassicals in regarding whole individual people as the paradigm or reference cases of agents. This is an essential assumption underlying any campaign to bring aspects of human psychology into the foundations of economic theory — as opposed to simply conjoining aspects of economics and psychology when specifically studying individual human choice. Now, if some who have employed paradigm shifting rhetoric want at this point to say that the latter idea is all they ever had in mind to promote, then disagreement dissolves. As noted above, I do not aim to tighten membership in the club of economists so as to exile the students of individual choice to another province where they must call themselves psychologists; such rigidity about disciplinary boundaries is silly. However, I claim that we dissolve the alleged basis for suggesting that economics is in theoretical crisis or would benefit from a paradigm shift if we give up the idea that the paradigmatic economic agent is a whole adult person. I will argue that the postwar practice of, and the direction of theoretical and practical progress in, economics is such that economists should be seen as venturing away from base camp whenever they turn their attention to non-aggregate phenomena. The contemporary concept of the agent is primarily a theoretical construction that facilitates modeling of aggregate phenomena; and it does a better job of this then would an agent fleshed out according to the profile of the human being furnished by psychologists.

3 ANIMAL AGENTS

As explained in the previous section, the agent in postwar economic theory is an abstraction. There are no manifest folk entities onto which agents need numerically map. In neuroeconomics, neurons and groups of neurons may be agents. In development economics, agents are statistically relevant households. In much macroeconomics since the 1970s, entire populations of countries are modeled as if they reflected a single 'representative' agent. By contrast, as also described above, the agent of BE is not abstract: she (no longer gendered, as in Jevons's time) is a manifest, living, breathing animal. More specifically, she is a *social* animal with a complex, multi-part control system that is too decentralized to produce the relentless consistency of the agent as previously defined.

Behavioral economists and their supporters among psychologists, philosophers and others have lately been remarkably successful in convincing other economists that in modeling agents they been neglecting important empirical considerations, and should feel chastened by discoveries coming from cognitive science generally and cognitive neuroscience particularly [Camerer *et al.*, 2005]. To cite one example, as Rubinstein [2007, p. 247] says "[t]en years ago it was difficult to publish a paper in the *QJE* which included a 'present-bias' assumption. These days it is almost impossible to publish a paper in the same journal which ignores present-bias, let

alone one which criticizes the approach."

The discoveries that are supposed to chasten mainstream economists can be broadly sorted 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. All of these are taken to threaten the supposed 'dogma' of mainstream (typically called 'neoclassical' or 'Walrasian') economics that people are rational and self-interested. The findings in sets (1)–(3) directly undermine (attributed) assumptions about peoples' practical consistency. Set (4) is often emphasized as undermining assumptions about narrow self-interest. This is an assumption which, it is quite easy to show, few economists make outside of institutionally constrained settings that specifically justify it [Cox, 2004; Weibull, 2004]. However, to the extent that people's preferences drift with those they pick up from reference groups, this will further undermine intertemporal consistency. Of course, none of these putative discoveries undermine the standard model of economic agency unless it is supposed that the paradigmatic economic agent is a natural (including socially constructed) person.

Rebel flags would not be flying from the battlements of top journals if many economists did not find the call for self-chastening persuasive. In aiming to resist it, I owe an account of this disposition to be humbled. The main part of the explanation, I believe, lies in the simplified history of their discipline that most economists imbibe from textbooks. Philosophers, whose discipline largely *consists* in its history, are apt to under-appreciate the extent to which economists, like most scientists preoccupied with achieving strikes into new terrain rather than consolidation behind the lines, typically get by with shallow narratives about the development of their paradigms. Any history of economics that gathers all 'neoclassicals', from Jevons through Samuelson to Chicago, into a single relatively homogenous doctrine is bound to be a caricature. So then working economists, highly alert to what works and doesn't work in the practice of modeling, can be readily brought to admit that the caricatured picture needs a fundamental makeover if they are to have a conceptual and methodological framework that is truly adequate to their knowledge and judgment. In addition, in my experience, no small number of economists suffer from an analogue to post-colonial guilt over their discipline's perceived arrogance as self-nominated 'queen of the social sciences'. The less nuanced BE manifestos tend to have a populist air; allowing that psychology might partly re-write basic economic theory is an obvious way to send a clear signal that economists have put imperialism behind them.

In the simplified history of thought that often frames casual (and some not-so-casual) methodological reflections in economics, it is acknowledged that economists

have a long history of ignoring psychologists. This, it is then frequently supposed, has stemmed from a conviction on economists' part that, in regarding people as narrowly selfish and materially motivated, they operated with a more realistic understanding of at least the *rational* parts of behavior than psychologists. But now, it is thought, BE empirically vindicates the psychologists, while still allowing an indispensable role for economists because of their training in formal modeling. In embracing the call for paradigm change inspired by BE, then, economists can refute the charge that their minds are closed to theoretical change motivated empirically and by non-economists, particularly the oft and unfairly neglected psychologists.

This impressionistic history of interdisciplinary relations isn't entirely false, of course; economists *do* have an established tradition of distancing themselves from psychology. As alluded to in the previous section, in the late 1930s and 1940s two threads in economic theory that had been developing separately were tied together. One thread was Keynes's focus on aggregate structural features of large economies without regard to the kinds of individual agents or actions that compose them[6] — that is, the then-new macroeconomics. The other was the attempt, clearly set in play by second-generation neoclassicists (Pareto and Fisher) near the turn of the century, to squeeze the psychological assumptions about economic agents down to a minimal core — ultimately, to *nothing but* consistency of preference rankings plus the idea that no agent would be content to consume only one type of good, no matter how cheap it became ('non-monomania'). Note that the second assumption is a substantial psychological hypothesis, and much more plausibly true of human beings than the first. Then, with Samuelson, as we saw, the need for even this final plausible human property was eliminated; we don't need to hypothesize non-monomania if we can use properties of observed demand to yield downward-sloping marginal utility functions empirically. This has frequently been interpreted, following the lead of Robbins [1935; 1938], as at last making a clean break between economics and psychology.

Despite their shared rejection of interpersonal comparisons of utility as unscientific, there is an important difference between the attitudes of Robbins and Samuelson toward scientific psychology. Whereas Robbins rejected the behaviorism then prevailing in psychology,[7] revealed preference theorists considered it to be a virtue of RPT — albeit, as I said earlier, a secondary one — that it was consistent with the up-to-date psychology of their time. They thus took it that ideas about how people internally represent their own preferences — most importantly for previous economists, their supposedly not subjectively liking each additional increment to their stock of a good as much as they subjectively liked the previous increment — are unscientific claims not just as economics but *as psychology.*

[6]Keynes is sometimes cited (e.g., by [Angner and Loewenstein, this volume]) as a precursor to psychologistic economics because he attributed business cycles to contagious emotions. However, this suggestion plays no direct role in his theory, which requires only that high-unemployment states be disequilibria. As later economists made much of, it *is* important that his theory assumes incomplete expectations on the part of consumers, producers and investors. But this was more of an oversight than an insight.

[7]He referred to it as a "queer cult" [Robbins, 1935, p. 87].

This point can be used to smooth the narrative that supports the self-chastening attitude. One can say: at least some important postwar economists *meant* to remain responsible members of a partnership with psychology, but then the profession missed the bus at the cognitive revolution in the 1960s. Fortunately, the paradigm shifters can continue, thanks to findings in experimental economics, to the undermining of aggregate welfare measures by Sonnenschein-Mantel-Debreu, and to the way in which game theory evolved, the bus eventually came around again and economists could redeem the earlier error by this time climbing aboard.

In Section 2 I referred to the fact that the rise of the refinement program in game theory plunged economists deep into modeling of belief profiles and other objects conceptualized using the language of psychological states. This encouraged interpretations of agency consistent with A $\Leftrightarrow$ O. But it simultaneously introduced a tension into this commitment by inflating the computational demands on agents. The players of many refined games — e.g., those that find so-called 'sequential equilibria' [Kreps and Wilson, 1982] — are computational prodigies, instantly updating all their beliefs, using all valid principles of Bayesian probability, upon receipt of any information. The capacities such refinements imply for agents are not plausible capacities of finite human beings whose inboard computational hardware was built by natural selection's incremental tinkering. And, sure enough, experimental economists duly showed that when people play the games analyzed by game theorists in laboratories, they often do not appear to behave like the agents in the models and they converge on vectors of strategies that are often not Nash equilibria (let alone subgame-perfect or sequential equilibria) according to the models [Camerer, 2003]. Thus, it seems to paradigm shifters, the 'assumptions' about agency of standard microeconomics need correction by the empirical facts of cognitive science.

The correction in question, according to the revolutionary manifestos, turns out to be drastic. People approximate traditional economic agency *behaviorally* in that they often accomplish their projects at bearable costs; but they don't exhibit *any* of the core *computational* properties attributed to economic agents by general equilibrium theory, rational-expectations macroeconomics, or game theory with refinement. 'Their' behavioral rationality typically turns out to really be natural selection's rationality, evolution having supposedly built rough situational rules of thumb ('heuristics') into people that serve them well as long as their environments are not too strange by comparison with their ancestral ones [Gigerenzer *et al.*, 1999]. This critique then appears to be reinforced by cognitive neuroscience, which musters evidence for biases, heuristics and framing effects operating directly in the processing systems of the brain [Camerer *et al.*, 2005]. Thus, it is concluded, economics collapses not just into abstract computational psychology, but all the way into computational neuroscience. That the word 'collapse' is not too strong is indicated by the sorts of things some neuroeconomists claim to discover. Recently, a team reported having determined from inspection of dopamine neurons that people do not value rewards by reference to their opportunity cost [Knutson *et al.*, 2007]; they infer from this that economic theory requires revision. Open-

ness to chastening from extra-disciplinary sources has gone remarkably far for any economist who admits that studies of the brain might imply revision in her view of opportunity cost as the basic state variable in microeconomics.

Once economics is taken to collapse into psychology, then discoveries in sets (1)–(4) above are naturally interpreted as tearing its standard theory apart. Furthermore, the news seems to have been getting worse since the early days of BE. Findings in sets (1)–(3) can, at least in principle, be accommodated by constructing new kinds of valuation functions. For example, people / agents can be taken to maximize *within* frames, even if not across them. Hyperbolic discount curves can be approximated by composing exponential ones of different slopes [Laibson, 1997; 1998]. However, cognitive science has lately been shaking free of a hyper-rationalistic and atomistic legacy of its own. The past decade has seen enormous upgrading of the significance attached to affect in explaining both mentation and behavior in people [Damasio, 1994; Panksepp, 1998]. Furthermore, affect itself is increasingly understood as both responding to and conditioning dynamic social interaction, an approach to modeling that seems to be borne out by the discovery of mirror neurons [Frith and Wolpert, 2004]. As individual people appear less and less to be autonomous bearers and computers of valuations, whose preferences explain their exchanges but are unchanged by them, and come instead to be seen as resembling adaptive nodes in social colonies where valuations continuously modulate one another in interacting cascades,[8] the more hopelessly inaccurate it is thought to be to model people, or aspects of them, as traditional economic agents.

Instead, it is suggested, the agent must cease to be 'bloodless'. This metaphor is apt, as we saw: the agent of classical economics (Sen's preferred model), and that of early neoclassicism, were not abstractions but organisms (or aspects of organisms). This will seem to be a banal observation if it is read simply as pointing out, with so many others, that BE aims to put emotions and lapses of rationality — failings of the flesh, as it were — back into economics. It is an equally familiar point that BE replaces the narrowly selfish agent with a socially concerned (both altruistic and envious) creature, though commentaries that make much of this often exaggerate, sometimes outrageously, the extent to which neoclassicism presupposes narrow selfishness. I want to emphasize something much less remarked upon in contrasting the (human) animal agent with the agent as characterized in Section 1. The former objects are, as it were, made by nature and 'found' in it by scientists, even if in modeling them they abstract away from all but a few of their properties; whereas the latter are not natural objects but constructed artifacts used to build models of phenomena that are, at least in the first place, social (in economists' jargon, either competitive or interactive/strategic).

Quite obviously, it could hardly be of greater importance or interest that we study the human organism. That study is, furthermore, sometimes crucial to applications of economic theory, especially when groups to which it is applied are small. However, I will now argue, study of the human organism is not *a part of* economics

[8]Such cascades are simulated in so-called 'swarm intelligence' models; see [Kennedy and Eberhart, 2001].

in a sense continuous with the core activity of postwar neoclassicism; whereas it is (of course) strongly continuous with psychology as practiced by Helmholtz and other founders of that discipline. There would be slightly less confusion abroad in the land, I suggest, if BE had instead been carried out under the label 'the psychology of valuation'. In saying this I am *not* asserting a normative claim about the 'proper' business of each discipline, or about how researchers ought to sort themselves among academic departments or about which journals should publish whose articles. On the contrary, I personally find it pleasing when the institutions of academe are allowed to become riots of methodological and conceptual diversity, at least insofar as this does not undermine the value attached to modeling rigor. Rather, what I mean to argue is that with respect to two substantively different scientific subject matters, which have historically been called 'psychology' and 'economics', BE is much more in the tradition of the former than the latter. Furthermore, BE no more implies that standard economic theory should undergo a revolutionary transformation than does any other part of psychology. I make this point by reference to ontology rather than methodology. BE, like psychology, studies the properties of people, whereas economics studies markets and networks, employing for this purpose an idea of 'agency' that is related to the concept of the person only by historical semantic tradition.

4 THE HEARTLAND OF ECONOMICS

To someone who both thinks that microeconomics is directly about individual human choice and behavior, and who also thinks that people are paradigmatic agents, the reason that agency is conceptually central to microeconomics needs little elaboration. As discussed in Section 2, if one is doubtful of the first two claims then the basis for the third is less obvious. In Section 2 I argued *that* agency indeed *is* central to microeconomics, given the sorts of modeling activities and analyses in which microeconomists in fact engage. However, I defended this claim strictly historically and pragmatically. Although I think that pragmatic considerations *are* highly relevant to ontology, I don't think that circumscribing the significance of philosophy should lead us to regard logic as irrelevant. The place of agency in economics should also partly be understood by reference to the logical structure of current theory.

The central objects of economic study are investment allocation, competition and strategic interaction. Economists investigate these processes by building models of their operations under different circumstances which are often, though not exclusively, inspired by real institutional environments. It is something like an analytic truth that competition and interaction must go on amongst distinct units; the economic agent is then whatever turns out to be the most serviceable concept of the competing or interacting unit. What mainly constrains this concept are features of the target explananda — which are, again, not the agents themselves, as in BE, but the competitive markets and interactive networks (which together largely determine the investment environment). Thus the properties of economic

agents, as captured in the analysis derived from Rubinstein in Section 1, are those that facilitate modeling of competition and strategic interaction.

A system is competitive (is a market) to the extent that agents have isomorphic utility functions and identical strategy sets given identical budget constraints. By 'isomorphic' I mean that if all goods are tradable and there is a fully fungible and liquid medium of exchange then agents can be modeled as if their utility functions differ only in index permutations: one utility function is designated as 'i's' and another as 'j's', where the claim that $i \neq j$ is primitive and entirely open to interpretation before a model is mapped onto an empirical subsystem of reality.[9] In a competitive setting, i and j aim at the same sort of end — e.g., maximization of expected monetary profits — except that i aims to maximize i's profits and j aims to maximize j's. If a market is *perfectly* competitive then, because no agents face special costs of capital or transaction costs, budget constraints are strictly functions of exogenous initial endowments and will converge if fluctuations in asset values are random walks. However, markets are imperfect if they include opportunities for earning rents or generating externalities, which may arise from asymmetries of information, from regulatory constraints, or from the existence of nonexcludable and/or non-rival goods. If agent i's utility maximization is constrained by j's maximizing behavior, then wherever these constraints are not fully captured by perfect market relationships i and j are members of an interactive network to be modeled as a game. Games in extensive form may be indefinitely embedded in one another, with terminal nodes of any one game assigned as initial nodes of others, and with payoff sets of outcomes expanded accordingly as agents are added by concatenation of new games. Since markets can be modeled without loss as games (trivial games in the case of perfectly competitive markets), game theory generalizes economics. This is important philosophically because it spares us any need to try to draw a crisp line between imperfectly competitive markets, systems that don't 'feel like' markets because many prices are shadow prices, and interactive networks where non-parametric factors dominate.

Which empirical substructures of models are identified by economists with agents is thus derivative on which empirical substructures they identify with markets and strategically interactive networks. The kinds of phenomena most often modeled as agents in economic applications are firms and households. In international economics, the agents are often countries. Typically, however, when firms, households and countries fail to behave as agents (e.g., exhibit cyclical preferences), we explain their behavior by 'breaking them up' into sub-agents, recognizing that CEOs and shareholders have different utility functions of their own, that treating husbands and their wives as unitary consumers often makes for misguided welfare policy, and that trade and exchange rate policies are temporary equilibria in dynamic games amongst producer lobbies and groups of politicians. Nevertheless, only in BE and in experimental economics are the phenomena identified with

[9]This phrase refers to standard model-theoretic semantic interpretation of scientific theory construction; see Ruttkamp [2002] for the formulation that, in my opinion, ideally equilibrates between explicitness and useful generality.

agents *usually* individual people.

This is of course not news to economists who nevertheless think that people are the paradigmatic agents (though I fear it *does* sometimes come as news to some philosophers who scarcely distinguish between microeconomics and decision theory). They may shrug it off precisely on the basis of emphasizing the previous point above. Even if the agents in most applied economics are aggregate, the standing pattern of disaggregating down to people when aggregate agency hits trouble shows that the *exemplary* agents still have the same identity they did for Jevons.

I think this is the strongest argument the advocate of $A \Leftrightarrow O$ has available. I am unpersuaded by it, however. A first part of the reason for this is a certain general view of the relationship between special sciences and philosophical ontology. I do not think that philosophers are entitled to suppose that where a science is inexplicit in practice about how its fundamental objects are related to those in other sciences or in metaphysics, philosophers perform a service when they infer the most parsimonious such set of relations they can and call this 'rational reconstruction'. This attitude rests on the idea that sciences have, as it were, background 'philosophical intentions' that transcend what their practitioners actually do, so that where scientific practice is silent or equivocal on metaphysics, philosophers may pipe up on its behalf. I don't see any evident justification for this attitude other than a very general belief that metaphysical commitment — any metaphysical commitment — is preferable to metaphysical agnosticism. And *that* belief, in turn, seems to me to have no justification at all.[10]

In light of this, I suggest that we should accept that economics is committed to $A \Leftrightarrow O$ *only* if we find applied economists actually making use of it in practice. A mere general tendency to decompose complex systems that exhibit imperfect agency into sub-agents falls short of this. What we would instead need to see is a working tendency to regard well-performing models in which the agents are individual people being regarded as authoritative over models in which the agents map onto some other sort of entity. Many readers will think this tendency is exhibited in economists' regularly manifest preference for models that can be given 'microfoundations'. Philosophers typically refer by microfoundations *either* to grounding compatible with an atomistic or individualist ontology *or* with grounding explanations in distinct physical objects with well-behaved boundaries (such as people) and concrete causal mechanisms (such as supposedly 'realistic' computations in people's brains). Economists generally mean by microfoundations something much more specific and *sui generis*: equilibria among sets of optimization functions. This is indeed a preference for agent-based models (and thus for interpretation (2) over (1)). Philosophers are apt to think that this economists' preference is merely a specific expression of *their* preference for decompositional reduction because they take for granted, contrary to what I have been arguing — and begging the question with respect to what is presently at issue — that a

[10]Davies [2009] argues that most contemporary philosophy is infected to its core with residues of theology. I agree.

preference for agent-based models necessarily indicates a commitment to $A \Leftrightarrow O$.

The hasty assumption I attribute to some philosophers readily arises from supposing that agent-based explanations get their 'grounding' (or at least purport to get it) from the idea that agents represent the targets of their optimizing behavior as goals, and that their 'rationality' consists in their literally computing plans of action to realize them. I do not doubt that organisms with brains represent and compute (though I certainly do doubt, following Clark [1997],[11] that representation and computation of the sorts of abstract relationships studied by economists are carried out entirely 'in organisms' heads'). However, what is important about agency for economists is consistent correlation of agents' behavioral responses with changes in relative scarcities (and hence in imputed opportunity costs), not — at least before the coming of the refinement program in GT — to any putative mechanistic basis for such responses. On a sufficiently abstract conception of computation, all responses to changes in relative scarcities are computed. But starfish, which are perfectly respectable agents, do not perform the relevant computations with their brains, because they do not have brains; dynamical coupling between naturally selected dispositions in their motor systems and environmental contingencies 'realize' the computations (as cognitive scientists say) and lead them to pursue prey and flee from predators in highly rational ways. A similar point can be made about a large firm: that strategy X, distributed over the aggregated behavioral tendencies of many branch offices, tends to maximize profits (or something else, like share value) in response to changes in supply or demand parameters does *not* entail that *any* individual person's brain, or any individual machine consulted by a person, explicitly represented or computed the relevant relationships. They may instead by stabilized by environmental constraints that no agents directly represent [Satz and Ferejohn, 1994].

Becker [1962] shows 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. Thus even the majority of applications in the area of economics most directly related in principle to the theory of choice, consumer theory, make no necessary working use of the supposed identity of economic agents and biological / psychological people. This fact should be taken at least as seriously as anything said about 'individual consumers' in opening chapters of introductory micro texts. I claim that a practical, philosophically fuzzy-minded, attitude about whether they are committed to a view on $A \Leftrightarrow O$ is what most economists *prefer* to any more explicit thesis that the philosophically motivated attempt to thrust upon them. Any claim to the effect that such a preference is feckless because metaphysical completion is a virtue of a scientific theory begs the question at issue. Pressed on the issue of just what their agents are, economists are quite entitled to say: anything in an empirical substructure of a model that, interpreted in light of the analysis of agency given in Section 1, yields predictive leverage and

[11] See especially Chapter 11.

explanation through integration with other established models.

Obviously, though, a significant number of important economists — BE polemicists, Sen, others — do *not* say this. Behavioral and experimental economists who resist this claim in a non-question-begging way (i.e., do not merely assume its denial in regarding their activity as economics instead of as psychology) may appeal to empirical discoveries about the way in which the brain computes reward values. I will deal with this basis for defense of A ⇔ O in the next section. For the moment let us remain in the heartland where neuroeconomic exotica are still unremarked. There, the two main developments in postwar theory discussed in Section 2 that blocked Samuelsonian elimination of agency from microeconomics altogether (the emergence of the refinement program in game theory and the attempt to derive welfare implications from general equilibrium theory) are sometimes conjoined with a largely *thoughtless* assumption of $A \Leftrightarrow O$ that is merely inherited from earlier neoclassicism. As a residual, philosophical, commitment to $A \Leftrightarrow O$, this is *not* what I have in mind by a *practical, working* commitment to it — a commitment that influences applied modeling.

When philosophers talk about 'practice' in a science they generally mean to refer to experimental protocols and accepted standards of evidence. This is still somewhat closer to epistemological norms than what I have in mind by 'practice' when considering a discipline that is as driven by engineering concerns as economics. Just as the de-psychologization of economics began before Samuelson, so did its increasing concentration on policy guidance, which in turn led to steady improvement in techniques for measuring and studying relations among aggregate variables — relations that are, or are at least widely thought to be, under the control of governments and central banks. The Keynesian revolution of the 1930s was an overnight triumph among economists because, as I mentioned in Section 3, in abandoning microeconomic modeling of macroeconomic phenomena Keynes was perceived as *liberating* the profession, exploiting his status as an all-around intellectual to give his more diffident colleagues license to dismiss ontological scruples they had maintained in deference to philosophical tradition. In the everyday practice of economics, *despite* the excitement over microfoundations that arose in the 1970s, there has been no looking back on this liberation. The overwhelming majority of working economists never estimate the utility function of an individual person. They measure elasticity coefficients of aggregate demand and production functions from changes in prices, interest rates, income distributions, national savings rates, and other index quantities. Most applied economists pay lip service to the idea that all of these things somehow 'boil down to' decisions by individual people. But by the weight of behavioral evidence this interest is usually perfunctory and the lip service is typically conventional. For example, textbooks in international economics admit that so-called 'community indifference curves' used to represent national welfare *cannot* be disaggregated into individual indifference curves without destroying the point of using them; most books cheerfully note this as a cautionary note and move on without further ado, assuming that the idea of 'national welfare' makes sense in its own right.

This is not to deny the clear fact that much economic theorizing in the mid-range between foundation building and specific applications consists in constructing microfoundations for models of aggregate-scale phenomena. However, the 'micro' here refers to the distinctive explanatory *logic* of microeconomic *theory*, not to decomposition of markets or networks into atoms. Let us consider an example. Going back to Tinbergen [1962], economists have represented trade flows between pairs of countries using so-called 'gravity models'. The original version of the gravity equation takes the form

$$M_{ij} = \alpha_k Y_i^{\beta_k} Y_j^{\gamma_k} N_i^{\zeta_k} N_j^{\upsilon_k} D_{ij}^{\sigma_k} \cup_{ijk}$$

where M_{ijk} is the value of the flow of good or factor k from country i to country j, Y_i and Y_j are income in country i and j respectively, N_i and N_j are populations of countries i and j, D_{ij} is the distance between countries i and j, and U_{ij} is a lognormally distributed error term with $E(U_{ijk}) = 0$ [Baldwin and Taglioni, 2006]. The name 'gravity model' derives from the fact that the equation represents a 'strength of attraction' based on countries' relative sizes and distances. Its original basis was intuitive and its justification in policy applications was for many years strictly empirical. It was not deemed fit to be regarded as a proper part of trade theory until it could be derived from a model of rational behavior by countries aiming to maximize returns on factors of production. An early effort by Anderson [1979], based on the assumption that goods produced in different countries are at best imperfect substitutes, was criticized for being *ad hoc* (but see [Anderson and van Wincoop, 2003; Baldwin and Taglioni, 2006]. More recently, Feenstra *et al.* [2001] proposed and empirically tested a microfoundational explanation widely thought to suffice, based on monopolistic competition that results from countries producing surplus differentiation of goods in consequence of optimizing inframarginal production efficiencies, and then engaging in mutually advantageous reciprocal dumping. Now, the point of this example in the present context is that among economists who think that Feenstra's account is empirically persuasive, it provides *sufficient* microfoundations for the gravity model because it shows why rational agents, which in this case happen to be countries, would produce and trade in accordance with the model's description. There is no further methodological requirement that the countries be disaggregated so that production of differentiated output can be attributed to particular models of firms; it is enough that the trade behavior optimizes inframarginal efficiencies and is a self-enforcing equilibrium. Thus 'microfoundations' here, as generally, refers not to ontological 'grounding out' in behavior of people as ultimate units, but to closing the model of an economic phenomenon in strictly *economic* terms, where 'economic' is defined by reference to an axiomatic theoretical system for identifying equilibria among behavioral dispositions or strategies of agents. Any requirement that these agents be individual people requires an extra-economic motivation.

Even in the realm of high theory, where microfoundations involve explanation by reference to agents learning to forecast monetary and fiscal policy, the agents in question are 'representative' optimizers whose ontological status is indeterminate.

In some canonical models whole economies are modeled as though they are single ('infinitely lived') agents whose business cycles result from the schedules on which they invest and take profits [Kydland and Prescott, 1982; Long and Plosser, 1983]. The underlying justification for this is the assumption that what are being modeled are markets in which utility functions differ only indexically. For this reason it doesn't *matter* to the formal analysis what sorts of extra-economic entities the utility functions map onto; all that matters is that econometric tests, based on measuring aggregate variables, can distinguish between one model and another. These tests require agents in the technical sense I have discussed; they do not require that the agents in question be people.

I advance a speculative counterfactual hypothesis about the sustaining motivation for concern with microfoundations in high theory. This speculation is that the devotion to constructing such foundations would not have been remotely as strong as it has been if the mathematics of microeconomic theory were not far more powerful and elegant than those of macroeconomics. Imagine for a moment a possible world in which this did not hold. In that world, if the mandarins of economic theory nevertheless put some of their best efforts into looking for microfoundations, this would have to be because they shared a driving philosophical conviction that sound explanations of phenomena must resemble those of an idealized version of classical physics, in which all principles boil down to mechanistic relationships among atoms. Mirowski [1989] argues persuasively that this was true of the early neoclassical economists; but, I contend, this is precisely the prison that Keynes unlocked. The possible world in which economic theorists are lashed forward by firmly maintained philosophical convictions seems very far from the one inhabited by actual current economists; an excellent way to persuade a typical economist to *drop* an opinion is to convince her it derives from a philosophical hunch. And there is in any case no pressing call to attribute philosophical faith to economists because a much more plausible account of the centrality of attention to microfoundations is readily available: economists want to deploy their most powerful technical toolkit, that of microeconomics, wherever they possibly can. This expresses a highly rational general principle. If application of a model of an infinitely lived representative agent allocating his future self-payments in an atomless measure space survives econometric testing then it would be foolish not to use the model in question. Infinitely lived agents and atomless measure spaces are hardly less *metaphysically* peculiar than flows of information and exchanged assets in complex systems that stabilize some such systems into markets. Metaphysical peculiarity or comfort simply have nothing to do with the matter.

Failure to appreciate that microfoundations means equilibrium dynamics rather than thoughts experienced by people has contributed to confused interpretations of what is politically and even morally at stake in macroeconomic policy debates. Consider, for example, the controversy between new classical macroeconomists and Keynesians over business cycles. Popular commentators frequently assert that the former show ideologically inspired callousness when they deny that there is 'involuntary' unemployment. However, as Lucas [1978] stresses in tones of justified

exasperation, a new classical theorist's microfoundational claim that all unemployment is voluntary is not about any aspect of any worker's psychological state, and thus does not possibly imply denial of the sincerity of anyone's misery or frustration; it is merely denial of the Keynesian claim that there are competitive equilibria in which human capital is wasted.[12] Microfoundational though it is, the macroeconomic dispute is about properties of markets, not about any properties of people.

So much for interest in microfoundations as a possible direct indicator of commitment to $A \Leftrightarrow O$. What about the possible indirect motivators identified earlier? In Section 2 I reviewed the two main developments in postwar economic theory that blocked Samuelsonian elimination of agency altogether. These were the emergence of the refinement program in game theory and the attempt to derive welfare implications from general equilibrium theory. Now I will say why I do not think that game theory provides a justified basis for doing economics according to the assumption that $A \Leftrightarrow O$. I will defer consideration of why we treat people as the proper objects of welfare concern to the very end of the essay.

In game theory, the refinement program largely expired by the turn of the century, mainly choking on a problem of its own rather than being smothered by the activities of economists turning into psychologists. The problem in question has a striking character in the present context: different possible refinements, applied separately or together, pulled economists' intuitions about rationality in conflicting directions. In consequence, game theory began increasingly to converge with, and become as unscientific as, the philosophy of ideal practical reason. Whether such philosophy is or is not a potential contributor to psychology — I here take no stand on that question — engagement with it has clearly seemed to most economists to be leading them away from their core business. The obvious way to reverse this drift into philosophy is the one that has mainly institutionally prevailed among economists: implement a stronger and cleaner distinction between 'rationality' in the thin sense — that is, Samuelsonian consistency of behavior with representation by preference orderings — and 'rationality' in the psychological sense of boundless in-board computational capacity.

In keeping with this, three main lines of research have taken centre stage among game theorists over the past ten years. One line applies classical game theory to contexts, such as auctions among highly capitalized players bidding for very valuable assets, in which institutional forces incentivize consortia to indeed behave like computational prodigies [Klemperer, 2004; Milgrom, 2004]. These consortia are not biological or psychological entities. Of course their representatives are such entities; but they are not imagined as doing their own computations, nor as choosing strategies using native, in-board cognitive resources. They have external computing equipment, including game theorist consultants with fancy software of their own. Second, game theorists have explored investment patterns in distributed markets by modeling them as games involving large numbers of players

[12]I do not intend here to imply preference for either side in this major and long-running theoretical controversy.

facing common uncertainty where all know that all know about the extent of uncertainty, and all know what technologies can be used to manage it (e.g., [Morris and Shin, 2003]). Here again is a use of game theory that eschews any appeal to psychological idiosyncrasies: players essentially use their models of the game situation to stabilize their expectations about one another, and they are embedded in institutional settings that are taken to constrain their utility functions, eliminating any special personal properties. Finally, the leading approach to multiple equilibria that has far overtaken appeal to refinements in popularity is application of evolutionary game theory [Weibull, 1995; Samuelson, 1998; Cressman, 2003]. This replaces the hyper-sophisticated agents of the refinement program with thoughtless players who simply inherit or copy strategies from others, with the probability of a strategy's getting inherited or copied being correlated with the strategy's success in previous rounds of iterated games. In this approach, strategies themselves, rather than agents, are the players of the games, with agents merely standing in to play their brief turns in a competitive process that continues beyond their individual lifespans. Agents must remain 'rational' in the thin sense — which is to say no more than that they remain agents — but much, most or all strategic and inferential computational demands are offloaded onto the selection process itself; thicker rationality 'goes virtual'. Young [1998] remains an exemplary set of applications.

Consideration of evolutionary game theory brings us to the edge of another kind of modeling that is rising in popularity in the more faddish precincts of economics, based on complex system theory [Anderson *et al.*, 1988; Arthur, 1994; Arthur *et al.*, 1997; Blume and Durlauf, 2005] It is noteworthy that many of the same people who advocate increased 'psychological realism' in economics are also fans of applying complex systems theory to social science (e.g. [Ormerod, 1999; Gintis, 2000; Beinhocker, 2006]). Denial of what philosophers call 'ontological reductionism'[13] — that is, atomism — is part of the very point of complex systems theory, with its emphasis on 'emergent' structures. These are properties and relations which are stabilized by bi-directional (that is, 'bottom-up' *plus* 'top-down) feedback relations and which cannot be decomposed into properties and relations of their parts. This new emergentism should, in my view, be approached with caution due to worries over stability of state variables across models. However, the simultaneous popularity, often in the same breasts, of extreme anti-reductionism *and* the view that economic theory ought to apply directly to individual objects with manifest boundaries is *prima facie* surprising. The odd conjunction suggests two things at once: tendencies in some quarters to favor ideas simply *because* they rebel against neoclassicism, and relatively reflexive assumption of $A \Leftrightarrow O$ that flies under theorists' radar because it is implicit, thereby sometimes capturing even those who are avowedly opposed to the intellectual tradition from which it is inherited.

[13]This locution is required to distinguish between reducing composite objects into parts, and reducing so-called 'high-level' theories to less abstract theories ('intertheoretical reduction'). Philosophers of science have generally been more interested in the latter than the former.

5 BEHAVIORAL ECONOMICS AND NEUROECONOMICS: THE MOLAR AND THE MOLECULAR

The argument of Section 4 was directed against interpreting recent trends in economic theory through the lens of ontological reductionism — more specifically, against interpreting economists' widespread interest in microfoundations as reflecting commitment to such reductionism. The most prominent current defenders of $A \Leftrightarrow O$ split into two camps in their attitudes to reductionism. Sen-style humanists oppose *psychological* reduction of O to A, preferring instead that $A \Leftrightarrow O$ be preserved by *inflating* A. Their motivations are largely grounded in normative considerations, upon which I will touch in my concluding remarks. Behavioral economists, by contrast, sometimes push for even more radical reductionism than is mandated by the $A \Leftrightarrow O$ thesis. Encountering violations of thin economic rationality in O-referenced behavior, they sometimes explain this by modeling people as corporate entities that emerge from the strategic interactions of sub-personal agents [Strotz, 1956].

I have elsewhere [Ross, 2005] argued for denial of $A \Leftrightarrow O$ from (as it were[14]) both 'below and above', and the idea that people are loci of — indeed are created and maintained by — strategic interaction of sub-personal agents is a concomitant of this denial that I have specifically endorsed and expanded upon [Ross *et al.*, 2008; Ross, 2009]. However, as part of the present essay's concern to resist the collapse of economics into psychology and/or neuroscience, I will here emphasize a tension within the decompositional approach. This arises over whether the sub-personal agents posited to explain economically relevant behavior of whole people are or are not identified with functional-anatomical parts of their brains.

In earlier work [Ross, 2005; 2006b] I have emphasized the contrast between *picoeconomics* and *neuroeconomics*. The term 'picoeconomics' was coined by Ainslie [1992; 2001] to denote applications of game theory to model what philosophers have traditionally called 'weakness of will' phenomena, including relapse to addiction, inconsistent financial saving, over-eating, and procrastination. Ainslie and other picoeconomists explain these common behavioral patterns as sometime equilibrium outcomes of games played amongst sub-personal interests, which arise as manifestations of hyperbolic discounting of future rewards at the personal scale. The identities of such interests are directly inferred from goals attributed at the personal scale by folk psychology. 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. The former interest might strengthen its prospects by 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 may give the smoking interest an advantage in short temporal ranges despite the fact that, from a longer range, the person's behavior reveals a preference for not smoking. (Typically, the most important such

[14]I add this locution to mark the fact that I elsewhere [Ladyman and Ross, 2007] am party to denial of the metaphysical image of reality as sorted into 'levels'.

behavior is voluntary suffering from restraint which would be pointless if relapse is sure; such behavior constitutes investment.) Whereas picoeconomics thus begins from the level of manifest behavior, neuroeconomics [Glimcher, 2003; Montague and Berns, 2002; Montague *et al.*, 2006] appeals to the ontology of anatomical and functional brain areas developed by neuroscience and identifies sub-personal agents, which may at times be in conflict, with functionally delineated groups of neurons (especially neurotransmitter systems). The utility functions of these units are implicit under a linear or dynamic programming interpretation of the algorithms they compute when physically healthy. Determination of these algorithms, mainly by comparing mathematical models with neuroimaging data, is the bread-and-butter work of the neuroeconomist.

People who are reluctant to acknowledge or have difficulty understanding the possible existence of anything that isn't a three (or four) dimensional hunk of matter [Heller, 1990] are apt to simply assume that if picoeconomic interests are not mere metaphors, they must ultimately reduce to neuroeconomic agents. However, this is inconsistent with Ainslie's understanding of the interests, which he identifies with their objects rather than their bearers. He is explicit that interests persist in time only for as long as the behavior they motivate is a standing possibility. Thus the procrastinator's interest in idly surfing the web while he tries to complete his tax return lasts for only as long as the task remains uncompleted or a less obviously unproductive distraction doesn't displace surfing in his attention. Of course, people have less fleeting interests such as in avoiding punishment or getting rebates from the government; willpower precisely consists in finding shorter-range interests that align with these, and by this device bringing the influence of the longer-range interests to bear on motivation in the present, where rewards are not hyperbolically discounted. Another of Ainslie's favorite examples is of an annoying interest in scratching an itch, which will fade entirely if even briefly ignored; unless the itch is caused by a foreign irritant, as most itches are not, the interest in scratching *is* the itch. Thus picoeconomic interests aren't sub-personal in the same sense as groups of neurons with specialist functions. The former are sub-personal in the sense that they have sharply limited projects that may not be endorsed by the whole person, but it is *molar* responses — behavior of a whole person at a time — with which they are associated. The agents of neuroeconomics, by contrast, are sub-personal in the sense of being molecular components of organisms.

The contrast between 'molar' and 'molecular' scales of description and explanation is a well established one in psychology, crucial to the behaviorist program from which picoeconomics descends. Molar-scale descriptions situate behavioral systems in environmental contexts, sorting their dispositions and properties by reference to equivalence classes of problems they face. These equivalence classes can 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.

By contrast, neuroeconomic models are computational and cognitivist in character. The 'economics' in neuroeconomics denotes a family of models of the way in which the so-called 'reward system' in the brain — roughly, the dopaminergic neurotransmitter system that projects from midbrain areas to orbitofrontal and pre-frontal cortex — comparatively values alternative allocations of attention, motor response and consumption. Such models provide algorithms by which the reward system is taken to estimate the expected opportunity costs of attending to one stimulus rather than another and of preparing one motor response rather than another. One of the current leading functional forms in the literature corresponds closely to the Black-Scholes model of portfolio option pricing [Montague and Berns, 2002]. In contrast to picoeconomic interests, which are often though not necessarily consciously accessible to people, neuroeconomic computational mechanisms never are. They are thus, to invoke a metaphor familiar to many economists, 'under-the-hood' causes of behavior. Psychologists refer to such trains of behavioral causation as 'molecular'. This talk is not intended to refer to chemistry, notwithstanding the importance of neurochemical agents to neuroeconomic applications. 'Molecular' here is intended purely as a logical contrast to 'molar', and is thus infrequent in the language of reductionists who deny the scientific validity of an autonomous molar scale.

Since molar-scale ontologies are developed by reference to organism-environment interfaces whereas molecular-scale ontologies are based on *in vitro* functions of internal computational organs, as a matter of logic molar and molecular scale models of one and the same system can vary independently. Of course logic cannot establish that they in fact *do* so vary, since this is an empirical matter. Strong reductionists expect that they don't, and thereby expect the molar scale to turn out to be redundant for psychological explanation. No one believes that they vary *completely* independently, since this would amount to denying that brains influence behavior.

Bearing in mind this contrast drawn in psychological terms, we can identify several different ways in which one might construct economic models of people and their behavior as reflecting interactions among sub-personal agents (or, in the case of the final alternative below, interactions between a unitary agent and non-agentic aspects of the organism):

(1A) One can model a person as *synchronically* composed of multiple sub-agents with conflicting utility functions (following the lead of Schelling [1978; 1980; 1984]. Then a pattern of personal-scale behavior might be modeled as the solution of a Nash bargaining game among these agents. (The restriction to Nash bargaining, as opposed to some other model of bargaining, might appear unmotivated. Note, however, that bargaining among synchronous sub-personal agents would have to be non-cooperative and un-governed by norms, lest the very point of so decomposing the person be lost. Under those assumptions Nash bargaining is the most general modeling framework.)

(1B) One can model a person as synchronically composed of multiple sub-agents with different time preferences. The reconstruction of hyperbolic personal time preference as resulting from competition between steeply exponentially discounting 'limbic'[15] regions and more patient (less steeply exponentially discounting) 'cognitive' regions [McClure *et al.*, 2004] is currently very popular with behavioral economists. In this kind of model, molecular-scale discounting with properties familiar to microeconomists is taken to explain molar-scale discounting featuring the properties emphasized by psychologists and behavioral economists.

(2) One can model a person as *diachronically* composed of multiple selves (each one of which controls the whole of a person's behavior for an interval of microseconds to hours) with differing utility functions and imperfect knowledge of one another, but where later agents' utility depends on investments by earlier agents. Then a pattern of personal behavior can be modeled as the subgame-perfect or sequential equilibrium of an extensive form signaling game in which agents choose actions with attention to the information this reveals about the probable preferences of their successors [Prelec and Bodner, 2003]. Since this has the effect of attaching some present utility to future rewards, it can (though of course it might not) implement willpower and correct for personal-scale intertemporal preference reversals that may otherwise arise due to hyperbolic discounting. Benabou and Tirole [2003] show in a full modeling exercise that such games can rationalize many of the suite of core picoeconomic behavioral phenomena described by Ainslie [1992; 2001] (but not one of his core explanatory targets, so-called reward building). These models of molar-scale phenomena involve no molecular-scale hypotheses at all.

(3) One can push the agentic aspect of the person 'deeper into the organism', in effect 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 and Pesendorfer [2001; 2005]. 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 simply a function of

[15]For years it was standard practice to refer to the older structure as the 'limbic system' and the newer brain as the 'cognitive system', based on the idea that emotional responses are primitive and rational ones are an adaptive refinement. As Paul Glimcher urges me to point out, over the past decade or so it has become clear that this is misleading; the older part of the brain performs many 'rational' calculations, and emotional judgments and motivations are crucial to the functioning of frontal cortex. However, it remains true that the older and newer parts of the brain developed under different evolutionary pressures.

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 quasi-hyperbolic curve. This sort of account straddles the molar/molecular divide, in describing and explaining rational behavior at the molar scale while explaining inconsistent consumption episodes by appeal to hypothesized molecular-scale disturbances. If this seems to reflect conflicted intuitions, a moment's reflection should render the source of the tension familiar: it simply amounts to keeping economics and psychology strongly separate. Agents remain abstract constructs, but humans in manifesting agent-like behavior are constrained by properties of their bodies. Interestingly, models of type (3) separate economics and psychology along the opposite polarity from Jevons, according to whom the economic aspects of the person pursue creature comforts while the psychological aspect can set its sights on nobler objectives.

Note that these three modeling approaches all reject $A \Leftrightarrow O$ in the *strict* sense (i.e., as analytic rather than as identification of a prototype; see Section 2), but in quite different spirits. Approaches 1A and 1B simply add isomorphic complexity to both sides of the equivalence so as to yield the following sort of picture:

$$\begin{aligned} &A_1 \Leftrightarrow O_1 \\ &A_2 \Leftrightarrow O_2 \\ &A_3 \Leftrightarrow O_3 \\ &\ldots \\ &A_n \Leftrightarrow O_n \end{aligned}$$

where $A_1, \ldots, A_n$ compose the agent $\mathbf{A}, O_1, \ldots, O_n$ compose the (brain of) the organism $\mathbf{O}$ and $\mathbf{A}$ and $\mathbf{O}$ are coextensive.

Approach (3) continues to numerically associate each basic agent with exactly one person, while allowing that the agent is only an aspect of the person. Approach (2) makes the person a derivative and sometime agent; a person achieves agency in the limited and temporary sense that a firm or country might, to the extent that intrapersonal signaling remains on an equilibrium path.

I will offer some provisional assessment of the relative current returns being delivered by these modeling strategies. Let the reader bear in mind here that it is still very early days for neuroeconomics and even the near future may not much resemble the immediate past.

Models of type 1 are certainly the most popular with neuroeconomic researchers. This is natural: science always tries to get as far as possible with reductionist models because they are conceptually, ontologically and structurally simplest. Indeed, we typically arrive at more complex models in science only through processes of correcting first-generation reductionist ones that turn out to be too simple in revealingly specific ways. An example of a type 1B neuroeconomic model could be obtained by setting the model of the dopamine reward system proposed by Schultz [2002] in the black box of the steep 'limbic' discounter (the 'β discounter') of McClure *et al.* [2004] and developing a correspondingly detailed model of their more patient 'cognitive' discounter (the 'δ discounter') to go along with it. This

example — the closest to a worked out one I am aware of — leads directly to an early intimation of the usual fate of straightforwardly reductionist models in our complex world: Glimcher *et al.* [2007] and Kable and Glimcher [2007] recently report fMRI data that they take to confute the hypothesis that different parts of the brain discount future rewards at different rates. The easier testability of reductionist accounts is their noble but tragic Popperian virtue.

It is important to point out here that models of the 1A type do not *have to* be read in a reductionist light. Suppose that, following Glimcher [2003], we interpret groups of neurons as economic agents. Suppose in particular that we so interpret the dopamine reward system. But now suppose that instead of reading the computational *processing* account of that system directly as the *economic* model of it, we derive its utility function by asking what its output would be if it optimized consistently given a maximally powerful statistical representation of its input data. (That is, suppose that we modeled it axiomatically instead of inductively.) This applies the concept of economic agency to the dopamine system in the same way that (non-behavioral) economists apply the concept to firms and households. In effect, it takes the economic model of the system to be a molar-scale account of the system in isolation, with a first-order computational account such as that of Schultz [2002] being its comparatively molecular counterpart processing model. (An account at the scale of cellular mechanisms would, on this picture, be comparatively molecular relative to the first-order computational one.) In light of the genesis and long history of the molar / molecular distinction in the stricter precincts of behaviorism, where all peeking under hoods was discouraged, this suggestion that there could be a *molar* account of a part of the brain is apt to seem strange and disorienting. However, it is not merely speculative. Recently, Caplin and Dean [2008] have furnished the first 'molar economic' model of the dopamine system *in vitro*. This model could in principle be used (for example) as input to an account of personal addictive behavior by setting it into a dynamic bargaining game with the correspondingly modeled inhibitory serotonergic system as its opponent, yielding a molar-scale economic complement to some currently popular molecular-scale neuropsychological accounts of addictive processes. The value of the economic model would lie in its potential identification of consumption properties that addiction might share with other, molecularly distinct, pathologies of impulsivity, which in turn could be expected to be relevant to policy and to non-pharmacological modes of treatment. See Ross *et al.* [2008] for more details of this picture. If this nascent approach to modeling bears empirical fruit, it should undermine the 'rebel' spin currently attached to BE about as directly as can be imagined, since it will preserve the separateness of economics from psychology in the exact Paretian spirit, while at the same time equally clearly violating $A \Leftrightarrow O$ 'from below'. I refer to this possible explanatory/modeling strategy as 'nerocellular economics', in recognition of the way in which it involves conceiving of sub-personal, functionally individuated agents as both neurally implemented in specifiable ways *and* as relatively autonomous optimizers from the modeling point of view.

Next let us consider type 3 models. In general, but again emphasizing the caveat about early days, models of this type are performing well in confrontation with data [Green and Myerson, 2004]. In light of the ontological flexibility of type 3 models, in which factors influencing behavior can be sorted pragmatically into exogenous and endogenous as suits the modeler, this is not surprising; while type 3 models often make excellent experimental design tools, Popperian virtues are not among those they parade. In this respect, type 3 models will have a familiar quality for both the economist and the most common kind of philosophical critic of economics (e.g. [Rosenberg, 1992]). I think it is a safe prediction that, given economists' strong interest in engineering applications — which, in the picoeconomic and neuroeconomic domains are mainly (potential) medical applications — type 3 models will be the most frequently observed over the coming years, even if modular neuroeconomic accounts sweep the boards with respect to unifying power, explanatory generality and theoretical rigor. Note, however, that because type 3 modeling rests on taking a casual attitude to ontological commitment, successes of such models *cannot* be used to establish that economics is a mere supplementary representational language for neuropsychology (cf. [Camerer *et al.*, 2005]) unless no less relaxed modeling strategies succeed and yield progressively improving track records. Existing type 3 models draw the distinction between agentic and non-agentic aspects of brain function in a way that is essentially arbitrary: why is a typical person's urge to slop cardiovascularly disastrous butter on her toast not an expression of her preferences while her standing attraction to a sports car, for which she might save for years, *is* such an expression? Gul and Pesendorfer [2001; 2005] define an exogenous temptation as a choice option for an agent with the property that its presence in the choice set makes the agent worse off, either because this results in her making a worse choice than she would have made in the option's absence, or because to cope with the option the agent must incur a cost of 'self-control'. This basis for distinction is clear enough for their operational purposes. But its only *justification* is pragmatic: it allows us to go on applying standard consumer theory in the face of apparent hyperbolic discounting and preference reversal. Pragmatism is a thoroughly respectable motivation for any economist; but it should not be expected to reveal unifying ontological principles — for example, that neuroscience describes 'real' processes to which economics should be expected to conform. (Gul and Pesendorfer agree.)

Finally, let us consider type 2 (picoeconomic) models. Scientists with reductionist intuitions are often inclined to regard them as beset by indeterminacies, and therefore as more like philosophical stories than scientific accounts. For example, should we expect a typical person's behavior to be described on the molar scale by one hyperbolic curve or many? Only the latter answer seems plausible. As Green and Myerson [2004] note, both temporally delayed and uncertain rewards are often discounted hyperbolically. However, people's degree of future discounting (their future-respective 'k-values', alluding to the standard equation[16]) are not good

[16] $v_i = A_i/(1+kD_i)$, where v_i, A_i, and D_i represent the present value of a delayed reward, the amount of a delayed reward, and the delay of the reward, respectively. The 1 in the denominator

general predictors of their uncertainty-respective k-values. Gambling addicts, for example, show the low relative concern for the future typical of all addicts (high future-respective k-vales) [Holt *et al.*, 2003], but also unusual tolerance for risk (low uncertainty-respective k-values) [Petry, 2001; Dixon *et al.*, 2003]. Ainslie [1992] observes that most people discount money less steeply than specific streams of consumption. Hoch and Loewenstein [1991] and Read [2001] point out that people do not hyperbolically discount future supplies of purely utilitarian (in their conceptual system, 'non-visceral') rewards such as petrol or computer paper; but we should not infer from this fact that they would not hyperbolically discount risk associated with the petrol supply. All of these points arise *despite* the fact that it is difficult to operationally disentangle intertemporal and uncertainty-based contingencies in economic models, since delay implies uncertainty outside of contexts where strict determinacy and perfect knowledge obtain, and (given instantaneous consumption) there can be no uncertainty about consumption without at least minimal delay. Finally, there is strong evidence that interval variance has some degree of influence on valuation of future rewards [Green and Myerson, 2004]; but, as Read [2001; 2003] objects, the picoeconomic framework abstracts away from this.

These indeterminacies would constitute embarrassments to picoeconomics only given a molecular interpretation of it. Ainslie and other advocates of picoeconomics (including me) have invited this interpretation by usually assuming that the picoeconomic model concerns delay discounting *rather than* probability discounting. This would invite a critic to suppose that the evidence of Glimcher, Kable and Louie [2007] and Glimcher and Kable [2007] mentioned earlier counterindicates the picoeconomic model along with its molecular-scale counterpart, the McClure *et al.* [2004] opponent brain-system model. A more careful interpretation of this evidence would have it as showing that the brain *does* implement computation of future discounting at a specific rate, while the behavioral phenomena discussed in the preceding paragraph are molar-scale generalities that hold *despite* the brain's discounting dispositions. Picoeconomic models should be regarded not as proto-neuroeconomic accounts of discounting, but as molar-scale profiles of the responses of organisms to differences in reward rates under different frames of attention. Exogenous influences from environments (including, in some organisms, social and cultural environments) likely play as critical a role in cueing and regulating these frames as do neural mechanisms. Thus we should not understand the picoeconomic agent as *composed out of* neuroeconomic ones.

The general conclusion I draw from these reflections is that there is room for all three types of models in the economics of personal and sub-personal behavior, though I am doubtful about the long-run viability of reductionist versions

prevents the rise in reward value from going infinite when delay is zero. The k parameter is a constant that is proportional to the degree of temporal discounting, with higher and lower k values describing greater and lesser degrees of discounting, respectively. Thus, an agent with a higher k value would discount delayed rewards more than an agent with a lower k value; the former agent therefore would be more impulsive than the latter.

of type 1 models. Apparent conflicts between picoeconomic and neuroeconomic approaches arise from assuming that there is a unique way of partitioning agents into sub-agents, so that a picoeconomic ontology of interests for a person must be isomorphic to a neuroeconomic ontology of brain areas for that person. The motivation for this is reductionism: the idea that molar-scale phenomena are in principle fully explicable by reference to molecular phenomena. But this is just a piece of philosophical dogma that fits the actual history of science very poorly [Ladyman and Ross, 2007]. The only empirically justifiable motivation for holding that one domain of modeling should reduce to another is actually observing the redundancy and abandonment, in that particular instance, of molar-scale models and their replacement by molecular-scale ones. I argued in earlier parts of the present essay that no such trend is manifest as between economics in general (i.e., outside of the avowed behavioral economics movement itself) and psychology or neuroscience. This does not at all imply that psychology and neuroscience are *irrelevant* to economics. The judgments of people, and of sub-personal picoeconomic interests, depend on neural computations of reward values as crucial input.; but neuroeconomics models the brain's valuations rather than the molar person's.[17] Thus (as in general) molecular-scale processes constrain molar-scale ones without reducing them.

The key implication of this form of anti-reductionism in the present context is that we can agree that people are not identical to economic agents without this necessarily implying that economic agency as traditionally understood is a useless or confused theoretical construct for explaining aspects of individual behavior. 'Necessarily' here needs emphasis. Rejecting an *a priori* motivation for collapsing economics into psychology does not in itself answer an obvious question implied in the criticism of standard microeconomics based on cognitive and behavioral science. That question is: if economic agents are asocial computational prodigies and people are constitutively social cognitive duffers, then what *is* the relationship between economic agents and people? To answer that there is *no* relationship *would* conjure up a mystery, except to a critic of mainstream economics so radical that she doubts that it ever succeeds at predicting anything.

I will argue in the concluding section of the chapter that, far from ignoring the social constitution of people, attention to this fact about them yields the answer to the question just posed.

6 PEOPLE AS COORDINATING EQUILIBRIA

One portentous claim emanating from the cognitive and behavioral sciences that is widely interpreted as implying trouble for mainstream economics is that people are pervasively, sub-consciously and irresistibly sensitive to manifold social cues, pressures and signals. Thus their preferences are not exogenous with respect to their

[17] For example, a group of dopamine neurons maximizes their utility by suppressing competing serotonergic circuits. If they are too successful the result is addiction, which is a disaster for the person and which few *people* want [Ross *et al.*, forthcoming].

strategic or consumption behavior. This claim lies at the core of Sen's [1977; 1999] critique of standard preference theory and what he calls 'welfarism'. A stronger claim is often made by anthropologists, sociologists and social psychologists that people are socially *constituted*. This claim is likely to strike many economists as a fundamental challenge to their way of thinking. However, in this final section of the chapter I will outline a perspective from which it is not. The basic idea is that once we get as far as recognizing people to be molar-scale objects[18] by comparison with their brains, then we can regard them as socially constituted without having to surrender the relevance of distinctively economic (as opposed to psychological) modeling to explanation of important aspects of their behavior. The perspective I will summarize here is not new, having been extensively elaborated in Ross [2005] and elsewhere. Readers are referred there for arguments. Here I will present, for the most part, only conclusions.

Human organisms are chemically integrated in meiosis, grown in the womb and then detached from their mothers' bodies at birth — they are not socially constructed. If it is nevertheless correct to claim that *people* are constituted socially, this must reflect the fact that they are *created* from human organisms by social development. Of course this process relies on properties of their brains: humans' giant cortex, and dispositions immanent in biases in neural connections and in the architecture of neurotransmitter pathways prepare them, unlike tigers, to be socialized. But the fact that we can distinguish between a very short pre-socialized phase and a socialized phase of a human organism's life supports a distinction between, as it were, the 'raw brain' and the person as a node in a dynamic social network. Raw human brains resemble tiger brains more than they resemble people. That people are socially constituted but their brains are not is the basic reason why behaviorists were right to emphasize the molar / molecular distinction. It doesn't suggest the dualist idea that persons *transcend* their brains; brains must adapt to socialization during development, and socialization is constrained by what brains can and cannot process.

To understand *how* people are socially created, something must first be said about why such developmental trajectories have been stabilized by selection. Let us distinguish between *social* animals and *herding* animals. Whereas the latter — wildebeest, for example, or corals — gain advantage merely by staying close together and coordinating their *schedules*, the former exploit efficiencies from joint contributions to ranges of projects that individuals can't perform alone, using some degree of specialization, either merely of talent or of dedicated roles. All available evidence suggests that natural selection, given the platforms it has had to work with in terrestrial history, can produce this in two ways: by adapting animals' genetic structures to increase the value of the inclusive coefficient in fitness functions, as in social insects and naked mole rats, or by adapting animals' brains so they develop enough book-keeping capacity to strategically discriminate among conspecifics and can thereby play strategic games involving reciprocal rewards and sanctions. High intelligence (cognitive plasticity) is far from continuously dis-

[18] In fact, people are better conceived as *processes* than as objects.

tributed across species, and sociality is far from continuously distributed across clades. It is thus of powerful significance under regression analyses that the entire hyper-intelligent club, which includes apes, elephants, dogs, toothed whales, corvids and parrots along with a few others, is social.

Within this club, humans are ecologically special in navigating an effectively boundless domain of novel collaborative projects. This is made possible by signaling systems — languages — that stabilize ranges of possible signal meanings by digitalizing information. That is, human syntax enables one human to direct another's attention to a specific object of reference even when it is not present to be pointed or gazed at; I can communicatively refer to 'Napoleon' exactly, not just to an indefinite range of things sharing to various degrees Napoleon's analog blend of properties (i.e., 'napoleonishness'). Thus humans can jointly track objects over time and space even when they are not present, and coordinate on future plans involving hypothetical objects picked out by digital contrast with other members of classes into which the grammars of public languages permit them to be sorted [Ross, 2007].

Some philosophers have suggested that language plus shared perceptual saliences are sufficient to account for people's ethologically unique capacity to coordinate. This is confused: the range of projects that can be distinguished thanks to recursive grammar makes the human coordination challenge orders of magnitude *more* complex than that faced by any other species. Game theorists encourage us to underestimate the difficulty of social coordination by solving for equilibria in situations they have already modeled as definite games. They readily forget that their own chief skill is in seeing how to abstract useful strategic models of empirical situations which don't come pre-packaged in terms of utility functions or strategy sets. Real human game players must implicitly construct models of their strategic situations in real time, without benefit of explicit principles, and they must jointly coordinate on these constructions; two interacting people who don't conceptualize their situation in terms of (roughly) the same game should expect not equilibrium but unpredictable chaos. Finally, let us bear in mind that every time a person takes an action she offers a move in a game with everyone whose welfare is potentially influenced by it and who might become aware of it — directly, by observing it or through gossip, or indirectly, by inferring it from outcomes, or second-order, by being influenced by the actions of someone else who is influenced by the original action. The overwhelming majority of human actions are thus simultaneously moves in multiple games with multiple sets of players of multiple n.

This all implies that most human choices of actions, no matter how small in scale, amount to general equilibrium problems. For example, to determine the best strategic response to my colleague's suggestion that we nominate a third colleague for a certain committee, I should, if I want to implement full rational agency, model the entire strategic history of our species (at least to the point in the future beyond which, due to discounting, I lose interest). This game is self-evidently intractable.

It gets still worse. A person's brain has a trillion neurons and 10^{13} synaptic connections, organized into semi-modular sub-systems that communicate imperfectly with one another, behave semi-autonomously and can no more be micro-managed by a frontal executive system than the President of the United States can plan every postal delivery and sentry assignment. These are of course the neuroeconomic agents discussed in the previous section. Not only do I not know the exact utility functions and strategy sets of the n other people with whom I'm strategically enmeshed, but I face significant uncertainty in predicting *my own* utility function and distribution of strategy sets, because much of my behavior is regulated by parts of my brain to which I have no more access than a third-person observer.

People clearly *do* coordinate, often very smoothly, over substantial stretches of time and place, and across large groups. Even more clearly, they don't do so by solving computationally impossible problems. The model of social coordination as solving for general equilibrium by solving an unbounded-n game *must* be missing something important. In social embeddedness and language, the very phenomena that lead to the impasse, lie the clues to what this something is. People sensibly insist that others with whom they enter into coordination games narrate comprehensible, publicly manifest stories about themselves and conform their behavior to these stories. Thus they enforce and enable predictability, including *self*-predictability. They mutually ease the imposed burden of this task by assisting each other as co-authors of narratives, recording expectations, rewarding enrichments of each other's sub-plots, and punishing overly abrupt attempts to revise important character dispositions. Parents initially impose this regime of self-construction on their children, later handing over primary control (often involuntarily) to their offspring's peer groups. Thus people become and remain distinct. The fact that self-creation and self-maintenance are *projects* requiring *effort* is what explains prevailing *normative* individualism, even while ('metaphysical') descriptive individualism is false. Individuals are centrally important to most of us partly *because* they don't just drop out of the womb. I will return to this point at the end of the chapter.

A crucial enabling aspect of this whole edifice is that humans are biologically adapted to be highly behaviorally sensitive to very cheap rewards (e.g. smiles, laughter, raised thumbs) and punishments (e.g. frowns, eye rolling, refusal of efforts at conversation). Not only are the standard punishments very inexpensive relative to the pain they inflict, but they can be withdrawn so as to leave almost no damaged infrastructure that then requires a new infusion of capital to put right; a person says "I forgive you" and the other's misery is (typically) instantly relieved. Some leading game theorists make the social coordination problem too hard, thereby motivating extravagantly hypothesized genetic adaptations to fix it, by exaggerating the costs of everyday rewards and punishments [Gintis, 2006; Seabright, 2006]. People avoid 'cheap talk' problems, in which their threats and promises would be ignored because it's doubted that these would be followed up if ineffective, by being psychologically adapted to care a great deal about rewards and punishments that cost others almost nothing [Ross, 2006a].

The effect of everyday pressures on people to construct and maintain selves is to drastically shrink the ranges of utility functions and strategy sets over which people must coordinate their constructions of games. The structures of these self-narratives then emerge as apparent framing effects and departures from proper Bayesian reasoning when we put people into experimental games and model these games as if the players weren't constrained by their own biographical and autobiographical plots.[19] This is a ubiquitous feature of the experimental literature in behavioral economics. Researchers define their subjects' games as if they were unconstrained by socialization, show that the outcomes do not match the Nash equilibria of these games, and thereby draw two generic conclusions (as background for various more specific conclusions that give us real psychological knowledge). The first sort is unobjectionable: people *are* constrained by socialization. But that is a truism, certainly known by Jevons, Walras, Samuelson, Milton Friedman and Robert Lucas alike. The second generic conclusion is that therefore standard economic theory is refuted because that theory is necessarily about unsocialized agents. This I reject.

I argued in previous sections that nothing in economic theory requires that economic agency be identified with individual people. Economic agency is a theoretical construction. Economists use it to build abstract models of firms, nations, labor unions, consortia in auctions, lineages in evolutionary games and other feedback-sensitive, incentive-driven systems that have no psychological properties at all. The usefulness of the construction is not cast into doubt by behavioral economics or by cognitive science more generally.

It is thus open to us to ask whether economics has *any* relevance to cognitive science (and hence to cognition understood as social). If the answer were 'no', economists in the spirit of Keynes might shrug this off and leave worries about unification of the sciences to philosophers. But the answer is not, in fact, negative. I just summarized an account of the universal human disposition to construct selves and to enforce such construction in one another. The explanation of this pattern is that it allows people to achieve many of the gains possible for economic agents — gains from trade, from specialization, and from consistent investment over time — despite the fact that their brains are too large and necessarily de-centralized as control structures to pull off economic agency by themselves. Thus economics plays a direct role in explaining the basis of social cognition. Furthermore, self-construction is only the first (necessary) aspect of the achievement of large-$n$coordination. The truly heavy lifting is done by the ultimate self-maintenance engines: institutions.

Most readers of this chapter will save money for relatively comfortable retirements. You will do this despite the fact that you would, if put in a systematically unfamiliar consumption environment, discount the future hyperbolically and therefore tend to reverse your preferences for prudent investments when temptations

[19]It's possible to induce people to escape from these constraints, in which case they tend to act much more like economic agents; but this requires deliberate effort in experimental design. See [Binmore, 2007].

to immediate reward presented themselves, then spend still more resources trying to defeat your own myopia as you learned the patterns governing the novel circumstances. Most of you will avoid this in your actual lives because your behavior is hemmed in and guarded by walls of culturally evolved and collectively designed institutions. If you persistently spend more than your income, this will be reflected in a falling credit rating that will inconvenience you *now*. Perhaps a recent housing bubble has allowed you to splurge for a few years, but as of this writing (mid-2007) market institutions are busy transmitting information about you and hundreds of millions like you that, through still other institutions, will correct your lack of prudence. If you aren't corrected quickly enough, the bank manager who supervises your mortgage may act to speed up receipt of the message. If very many of you are too sluggish responding to the news, the Chairman of the Federal Reserve Bank may reinforce it with an interest rate hike. And so on.

All of these institutions press you to approximate your behavior to that of an economic agent. They can't literally transform you, biological — psychological entity that you are, into such an agent. Even while struggling to save, you may visit a casino. You will buy some items this year that you will disdain and throw away in a year's time merely because your tastes change. But you, together with your fellows in society, have *enough* in common with economic agents, especially in modern institutional settings, that non-trivial predictions about your individual behavior can be had by modeling you as if, within temporal and institutional constraints, you were such agents. Furthermore, because you live in aggregated markets with dynamics that aren't very sensitive to psychological factors, *and* because you also play n-person games with other agents who are incentivized to stabilize one another's preference consistency, you can improve your prospects by learning some economic theory and feeding this social knowledge back into your personal planning. Feedback loops of this sort are the very logical essence of social cognition. *Both* your person-hood *and* your approximate economic agency — which, I have argued, are not the same thing — are socially constituted.

Individualism is thus descriptively false. As explained above, that is part of the reason *why* it is *normatively* important. This insight should allow us to see that we don't need to justify concerns for aggregate welfare by disaggregating it — which we can't in general do, as Arrow's theorem makes clear. The proper normative defense of macroeconomics without microfoundations has two parts, one familiar and narrowly economic and one less familiar and broader. First, if a policy takes a society to a higher community indifference curve than it was on before, but the new allocation and the old are Pareto-noncomparable, then we should still find that winners can compensate losers using less than the whole of their winnings; the new policy should bring about a Scitovsky-Kaldor-Hicks improvement. Second, we should see this as a *normative* improvement on utilitarian grounds *because* individual preferences are not exogenous. As modeled by Binmore [1998], people will bargain to a new distribution under the new dispensation and then they will adjust their distributive norms — that this, their collectively determined concept

of justice — so as to rationalize the bargaining outcome. This will not at all impress a philosopher with Kantian intuitions, since the result may fail to 'respect' any given person's prior idea of fairness — justice is de-coupled from individual autonomy. But under the perspective I have defended here, such autonomy is a myth anyway if regarded as meaningful *outside of* an institutional specification. Such a specification is a norm-governed network. (It will happen now and then to be a market. In these unusual circumstances norms of justice doesn't matter and are only applied when people get confused.) When people adjust their norms they approximate different agents.

The Kantian philosopher is unimpressed by this story because she doesn't see any touchstone against which to regard the distribution on the higher community indifference curve as necessarily *better*. But the economist has an evaluative standard: the people are materially richer. The economic agents they formerly approximated may or may not have all had their preferences optimized; this we can't tell, for both economic and philosophical reasons. The economic reason is that Scitovsky-Kaldor-Hicks improvements aren't necessarily Pareto-improvements. The philosophical reason is that non-autonomous agents before and after institutional norm-readjustment are different agents. But although economics studies such agents as its first-order objects, and although these agents are not identical to the more enduring human entities that approximate sequences of them, the ultimate *justification* of economics is that it is useful for guiding our efforts to make *material* human animals materially better off. In a world not merely of pervasive scarcity but much outright poverty, the justification for the philosophical ethicist's activities seems to me to be comparatively thin gruel.

Thus, I conclude, a defense of economics as both objective science and normatively helpful engineering is best articulated without $A \Leftrightarrow O$. Economics is not, and should not become, a kind or branch of psychology. It is about agents, in the sense that it is interactions of agents about which it makes discoveries; and the agents it is about are not people. Its discoveries are nevertheless very important to people.

BIBLIOGRAPHY

[Anderson, 1979] J. Anderson. A theoretical foundation for the gravity equation. *American Economic Review* 69: 106-116, 1979.

[Anderson and van Wincoop, 2003] J. Anderson and E. van Wincoop. Gravity with gravitas: a solution to the border puzzle. *American Economic Review* 93: 170-192, 2003.

[Anderson *et al.*, 1988] P. Anderson, K. Arrow, and D. Pines, eds. *The Economy as an Evolving Complex System.* Boston: Addison-Wesley, 1988.

[Anger and Loewenstein, 2010] E. Angner and G. Loewenstein. (this volume). Behavioral economics.

[Arrow and Debreu, 1954] K. Arrow and G. Debreu. Existence of equilibrium for a competitive economy. *Econometrica* 22: 265-290, 1954.

[Arthur, 1994] W. B. Arthur. *Increasing Returns and Path Dependence in the Economy.* Ann Arbor: University of Michigan Press, 1994.

[Aruthur *et al.*, 1997] W. B. Arthur, S. Durlauf, and D. Lane, eds. *The Economy as an Evolving Complex System II.* Boston: Addison-Wesley, 1997.

[Baldwin and Taglioni, 2006] R. Baldwin and D. Taglioni. Gravity for dummies and dummies for gravity equations. NBER Working Papers 12516, 2006: `http://ideas.repec.org/s/nbr/nberwo.html`

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

[Beinhocker, 2006] E. Beinhocker. *The Origins of Wealth.* Cambridge, MA: Harvard Business School Press, 2006.

[Benabou and Tirole, 2003] R. Benabou and J. Tirole. Willpower and personal rules. *Journal of Political Economy* 112: 848-886, 2003.

[Binmore, 1990] K. Binmore. *Essays on the Foundations of Game Theory.* Oxford: Blackwell, 1990.

[Binmore, 1998] K. Binmore. *Game Theory and the Social Contract, Volume Two: Just Playing.* Cambridge, MA: MIT Press, 1998.

[Binmore, 2007] K. Binmore. *Does Game Theory Work? The Bargaining Challenge.* Cambridge, MA: MIT Press, 2007.

[Blume and Durlauf, 2005] L. Blume and S. Durlauf, eds. *The Economy as an Evolving Complex System III.* Oxford: Oxford University Press, 2005.

[Bruni, 2005] L. Bruni. *Hic sunt leones*: interpersonal relations as unexplored territory in the tradition of economics. In B. Gui and R. Sugden (Eds.), *Economics and Social Interaction* (pp. 206-228). Cambridge: Cambridge University Press, 2005.

[Bruni and Sugden, 2007] L. Bruni and R. Sugden. The road not taken: how psychology was removed from economics and how it might be brought back. *The Economic Journal*, 117, 146-173, 2007.

[Camerer, 2003] C. Camerer. *Behavioral Game Theory.* Princeton: Princeton University Press, 2003.

[Camerer and Loewenstein, 2004] C. Camerer and G. Loewenstein. Behavioral economics: Past, present and future. In C. Camerer, G. Loewenstein and M. Rabin, eds., *Advances in Behavioral Economics*, pp. 3-51. Princeton: Princeton University Press, 2004.

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

[Caplin and Dean, 2008] A. Caplin and M. Dean. Dopamine and reward prediction error: an axiomatic approach to neuroeconomics. *American Economic Review*, 97: 248–152, 2008.

[Cox, 2004] J. Cox. How to identify trust and reciprocity. *Games and Economic Behavior* 46: 260-281, 2004.

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

[Damasio, 1994] A. Damasio. *Descartes's Error.* New York: Putnam, 1994.

[Davies, 2009] P. S. Davies. *Subjects of the World.* Chicago: University of Chicago Press, 2009.

[Davis, 2003] J. Davis. *The Theory of the Individual in Economics.* London: Routledge, 2003.

[Debreau, 1959] G. Debreu. *Theory of Value.* New York: Wiley, 1959.

[Debreu, 1960] G. Debreu. *Mathematical Methods in the Social Sciences.* Stanford: Stanford University Press, 1960.

[Dixon *et al.*, 2003] M. Dixon, J. Marley, and E. Jacobs. Delay discounting by pathological gamblers. *Journal of Applied Behavior Analysis* 36: 449–458, 2003.

[Feenstra *et al.*, 2001] R. Feenstra, J. Markusan, and A. Rose. Using the gravity equation to differentiate among alternative theories of trade. *Canadian Journal of Economics* 34: 430-447, 2001.

[Friedman, 1953] M. Friedman. *Essays in Positive Economics.* Chicago: University of Chicago Press, 1953.

[Frith and Wolpert, 2004] C. Frith and D. Wolpert, eds. *The Neuroscience of Social Interaction.* Oxford: Oxford University Press, 2004.

[Fullbrook, 2003] E. Fullbrook, ed. *The Crisis in Economics.* London: Routledge, 2003.

[Ghemawat, 1998] P. Ghemawat. *Games Businesses Play.* Cambridge, MA: MIT Press, 1998.

[Gigerenzer *et al.*, 1999] G. Gigerenzer, P. Todd, and the ABC Research Group. *Simple Heuristics that Make Us Smart.* Oxford: Oxford University Press, 1999.

[Gintis, 2006] H. Gintis. Behavioral ethics meets natural justice. *Politics, Philosophy and Economics* 5: 5-32, 2006.

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

[Glimcher *et al.*, 2007] P. Glimcher, J. Kable, and K. Louie. Neuroeconomic studies of impulsivity: now or just as soon as possible? *American Economic Review*, 97(2): 142–147, 2007.

[Green and Myerson, 2004] L. Green and J. Myerson. A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin* 130: 769 — 792, 2004.

[Gul an dPesendorfer, 2001] F. Gul and W. Pesendorfer. Temptation and self control. *Econometrica* 69: 1403-1436, 2001.

[Gul and Pesendorfer, 2005] F. Gul and W. Pesendorfer. The simple theory of temptation and self-control, 2005. `http://www.princeton.edu/~pesendor/finite.pdf`

[Heller, 1990] M. Heller. *The Ontology of Physical Objects.* Cambridge: Cambridge University Press, 1990.

[Heilbroner and Milberg, 1995] R. Heilbroner and W. Milberg. *The Crisis of Vision in Modern Economic Thought.* Cambridge: Cambridge University Press, 1995.

[Hoch and Loewenstein, 1991] S. Hoch and G. Loewenstein. Time-inconsistent preferences and consumer self-control. *Journal of Consumer Research* 17: 492-507, 1991.

[Hollis and Nell, 1975] M. Hollis and E. Nell. *Rational Economic Man.* Cambridge: Cambridge University Press, 1975.

[Holt *et al.*, 2003] D. Holt, L. Green, and J. Myerson. Is discounting impulsive? Evidence from temporal and probability discounting in gambling and non-gambling college students. *Behavioural Processes* 64: 355–367, 2003.

[Jevons, 1871] W. S. Jevons. *The Theory of Political Economy.* London: Macmillan, 1871.

[Kable and Glimcher, 2007] J. Kable and P. Glimcher. The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, 10: 1625-1633, 2007.

[Kennedy and Eberhart, 2001] J. Kennedy and R. Eberhart. *Swarm Intelligence.* San Fransisco: Morgan Kauffman, 2001.

[Klemperer, 2004] P. Klemperer. *Auctions: Theory and Practice.* Princeton: Princeton University Press, 2004.

[Knutson *et al.*, 2007] B. Knutson, S. Rick, G. E. Wimmer, D. Prelec, and G. Loewenstein. Neural predictors of purchases. *Neuron*, 53, 147-156, 2007.

[Kreps and Wilson, 1982] D. Kreps and R. Wilson. Sequential equilibrium. *Econometrica* 50: 863-894, 1982.

[Kydland and Prescott, 1982] F. Kydland and E. Prescott. Time to build and aggregate fluctuations. *Econometrica* 50: 1345-1369, 1982.

[Ladyman and Ross, 2007] J. Ladyman and D. Ross. *Every Thing Must Go.* Oxford: Oxford University Press, 2007.

[Laibson, 1997] D. Laibson. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112, 443-477, 1997.

[Laibson, 1998] D. Laibson. Life-cycle consumption and hyperbolic discount functions. *European Economic Review*, 42, 861-871, 1998.

[Lipsey and Lancaster, 1956] R. Lipsey and G. Lancaster. The general theory of second best. *Review of Economic Studies*, 24: 11-32, 1956.

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

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

[Long and Plosser, 1983] J. Long and C. Plosser. Real business cycles. *Journal of Political Economy* 91: 39-69, 1983.

[Lucas, 1978] R. Lucas. Unemployment policy. *American Economic Review* 68: 353-357, 1978.

[Mäki, 1986] U. Mäki. Rhetoric at the expense of coherence: a reinterpretation of Milton Friedman's methodology. In W. Samuels, ed., *Research in the History of Economic Thought and Methodology, Volume Four*, pp. 127-143. Greenwich, CT: JAI Press, 1986.

[Mäki, 1992] U. Mäki. Friedman and realism. In W. Samuels and J. Biddle, eds., *Research in the History of Economic Thought and Methodology, Volume Ten*, pp. 171-195. Greenwich, CT: JAI Press, 1992.

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

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

[Mandler, 1999] M. Mandler. *Dilemmas in Economic Theory.* Oxford: Oxford University Press, 1999.

[McClure *et al.*, 2004] S. McClure, D. Laibson, G. Loewenstein, and J. Cohen. Separate neural systems value immediate and delayed monetary rewards. *Science* 306: 503-507, 2004.

[Milgrom, 2004] P. Milgrom. *Putting Auction Theory to Work.* Cambridge: Cambridge University Press, 2004.

[Mirowski, 1989] P. Mirowski. *More Heat Than Light.* New York: Cambridge University Press, 1989.

[Mirowski, 2002] P. Mirowski. *Machine Dreams.* Cambridge: Cambridge University Press, 2002.

[Montague and Berns, 2002] P. R. Montague and G. Berns. Neural economics and the biological substrates of valuation. *Neuron* 36: 265-284, 2002.

[Montague *et al.*, 2006] P. R. Montague, B. King-Casas, and J. Cohen. Imaging valuation models in human choice. *Annual Review of Neuroscience* 29: 417-448, 2006.

[Morris and Shin, 2003] S. Morris and H.S. Shin. Global games: theory and applications. In M. Dewatripont, L.P. Hansen and S Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eight World Congress, Volume 1*, pp. 56-114, 2003.

[Ormerod, 1994] P. Ormerod. *The Death of Economics.* New York: Wiley, 1994.

[Panksepp, 1998] J. Panksepp. *Affective Neuroscience.* Oxford: Oxford University Press. 1998.

[Petry, 2001] N. Petry. Pathological gamblers, with and without substance abuse disorders, discount delayed rewards at high rates. *Journal of Abnormal Psychology* 110: 482-487, 2001.

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

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

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

[Robbins, 1935] L. Robbins. *An Essay on the Nature and Significance of Economic Science,* 2^{nd} edition. London: Macmillan, 1935.

[Robbins, 1938] L. Robbins. Interpersonal comparisons of utility: a comment. *Economic Journal* 43: 635-641, 1938.

[Rosenberg, 1992] A. Rosenberg. *Economics: Mathematical Politics or Science of Diminishing Returns?* Chicago: University of Chicago Press, 1992.

[Ross, 2002] D. Ross. Why people are atypical agents. *Philosophical Papers* 31: 87-116, 2002.

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

[Ross, 2006a] D. Ross. Evolutionary game theory and the normative theory of institutional design: Binmore and behavioral economics. *Politics, Philosophy and Economics* 5: 51-79, 2006.

[Ross, 2006b] D. Ross. The Economics of the sub-personal: two research programs. In B. Montero and M. White, eds. *Economics and the Mind*, pp. 41-57.London: Routledge, 2006.

[Ross, 2007] D. Ross. *H. sapiens* as ecologically special: what does language contribute? *Language Sciences*, 2007.

[Ross *et al.*, 2008] D. Ross, C. Sharp, R. Vuchinich, and D. Spurrett. *Midbrain Mutiny: The Picoeconomics and Neuroeconomics of Disordered Gambling.* Cambridge, MA: MIT Press, 2008.

[Ross, 2009] D. Ross. Integrating the dynamics of multiscale economic agency. In H. Kincaid and D. Ross, eds., *The Oxford Handbook of Philosophy of Economics*, pp. 245-279. Oxford: Oxford University Press, 2009.

[Rubinstein, 2006] A. Rubinstein. *Lecture Notes in Microeconomic Theory: The Economic Agent.* Princeton: Princeton University Press, 2006.

[Rubinstein, 2007] A. Rubinstein. Discussion of behavioral economics. In R. Blundell, W. Newey and T. Persson, eds., *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress*, pp. 246–254. Cambridge: Cambridge University Press, 2007.

[Ruttkamp, 2002] E. Ruttkamp. *A Model-Theoretic Realist Interpretation of Science.* Dordrecht: Kluwer, 2002.

[Samuelson, 1998] L. Samuelson. *Evolutionary Games and Equilibrium Selection.* Cambridge, MA: MIT Press, 1998.

[Samuelson, 1938] P. Samuelson. A note on the pure theory of consumer's behavior. *Economica* 5: 61-72, 1938.
[Samuelson, 1947] P. Samuelson. *Foundations of Economic Analysis.* Enlarged edition (1983). Cambridge, MA: Harvard University Press, 1947.
[Satz and Ferejohn, 1994] D. Satz and J. Ferejohn. Rational choice and social theory. *Journal of Philosophy* 91: 71-87, 1994.
[Schelling, 1978] T. Schelling. Economics, or the art of self-management. *American Economic Review* 68: 290-294, 1978.
[Schelling, 1980] T. Schelling. The intimate contest for self-command. *Public Interest* 60: 94-118, 1980.
[Schelling, 1984] T. Schelling. Self-command in practice, in policy, and in a theory of rational choice. *American Economic Review* 74: 1-11, 1984.
[Schultz, 2002] W. Schultz. Getting formal with dopamine and reward. *Neuron* 36: 241-263, 2002.
[Seabright, 2006] P. Seabright. The evolution of fairness norms: an essay on Ken Binmore's *Natural Justice. Politics, Philosophy and Economics* 5: 33-50, 2006.
[Sen, 1977] A. K. Sen. Rational fools. *Philosophy and Public Affairs* 6: 317-344, 1977.
[Sen, 1999] A. K. Sen. *Development as Freedom.* New York: Random House, 1999.
[Sonnenschein, 1972] H. Sonnenschein. Market excess demand functions. *Econometrica*, 40: 549-563, 1972.
[Sonnenschein, 1973] H. Sonnenschein. Do Walras identity and continuity characterize the class of excess demand functions? *Journal of Economic Theory*, 6: 345-354. 1973.
[Stigum, 1990] B. Stigum. *Toward a Formal Science of Economics.* Cambridge, MA: MIT Press, 1990.
[Strotz, 1956] R. Strotz. Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies* 23: 165-180, 1956.
[Tinbergen, 1962] J. Tinbergen. *Shaping the World Economy.* New York: The Twentieth Century Fund, 1962.
[van Fraassen, 1980] B. van Fraassen. *The Scientific Image.* Oxford: Oxford University Press, 1980.
[van Fraassen, 2002] B. van Fraassen. *The Empirical Stance.* New Haven: Yale University Press, 2002.
[Weibull, 1995] J. Weibull. *Evolutionary Game Theory.* Cambridge, MA: MIT Press, 1995.
[Weibull, 2004] J. Weibull. Testing game theory. In S. Huck, ed., *Advances in Understanding Strategic Behavior,* pp. 85-104. Houndmills, Basingstoke, Hampshire: Palgrave, 2004.
[Young, 1998] H. P. Young. *Individual Strategy and Social Structure.* Princeton: Princeton University Press, 1998.