7-1-2007

Review of *Economic Theory and Cognitive Science* by Don Ross

John B. Davis
Marquette University, john.davis@marquette.edu

Accepted version. *Economics and Philosophy*, Vol. 23, No. 2 (July 2007): 245-252. DOI.

a transitive and complete betterness relation that incorporates all relevant aspects or perspectives, there does not seem to be a plausible alternative to maximizing.

Krister Bykvist
Jesus College, Oxford

doi:10.1017/S026626710700140X

Economic Theory and Cognitive Science, by Don Ross. MIT Press, 2005, 384 pages.

Don Ross' *Economic Theory and Cognitive Science* is a challenging, well thought out book that exhibits considerable understanding of economics, philosophy, and cognitive science, and deserves to be taken seriously. Its premise that economics must not only address cognitive science but change in response to it seems entirely correct, though if current experience is any indication most economists will realize this only long after the die of change is cast. Of the many things worth discussing in the book, I will not address Ross' critique of eliminative materialism and intentional-stance functionalism alternative (cf. Nagel 1986; Searle 1997), his reading of the history of economics regarding Robbins and Samuelson, his view that reality boils down to a fundamental unity of one underlying kind of stuff (cf. Dupré 2001), his separateness of economics thesis (cf. Hausman 1992), nor his radical scientific realism and rejection of commonsense ontology (cf. Mäki 1992). I will address what I take to be the pivotal focus of the book, namely the ontological thesis that human individuals or selves are not agents but their subpersonal aspects are. I begin with a summary of Ross' relevant arguments, and then move to their evaluation.

RESCUING NEOCLASSICISM?

Ross seeks to unite the 'core insights of neoclassical economics with evolutionary cognitive and behavioral science' in a way that abandons both 'our conventional, "folk" schema for sorting intentional, behavioral and social reality' (19), and also the traditional assumption that human individuals or selves are agents. His approach is the opposite of that recommended by many other advocates of behavioral and evolutionary economics (e.g. Bowles 2003), who see the new programs as essentially anti-neoclassical. Ross rejects their position as relying on 'hyperempiricist methodological principles' and as a misguided attempt to transform economics into 'a branch of applied social psychology' (28), violating the purported status of economics as a separate science. Rather he argues that 'the core neoclassical commitment to economics as the systematic science

of maximization under scarcity comports *better* with the most sophisticated philosophy of cognitive science than does emphasis on unsystematic hyperempiricism' (29). Thus he offers an 'updating and defense of neoclassicism in the light of cognitive science', though allows that his picture will 'look strikingly different from the one usually associated with neoclassicism, and will jettison a number of theses to which neoclassicists are widely regarded as committed', including that individual people generally are rational maximizers of expected utility, individualism, and that people are generally selfish (28). Indeed evolutionary dynamics will carry greater weight in explaining economic behavior than any deliberate rational calculation carried out by individual economic agents.

Ross also rejects an internalist view of mental states and intentionality, and approaches propositional attitudes in a functionalist manner (rather than in an eliminative materialist way). The 'Dennettian package' or intentional-stance functionalism Ross relies on seeks to explain (not dissolve) consciousness and the self without appeal to intrinsic meaning by offering an account of propositional attitudes in terms of 'triangulated regularities among a subject, features of her environment, and patterns of expectations of her interpreters' (49). Thus, if in a traditional folk psychology manner one wants to attribute desires and beliefs to a given agent, then foregoing an internalist approach means one focuses on the 'network of social facts about language *and* a standing set of behavioral regularities' associated with agents (53). From a mereological perspective (seeing reality as a unity of one kind of underlying stuff), the issue for an intentional-stance functionalism is whether one can carry out a progressive mereological simplification (in the intertheoretic but not reductionist sense) that treats propositional attitudes as descriptions of patterns of social communication. Key here for Ross is Shannon's (1948) treatment of information in physically measurable terms, such that economics advances in tandem with cognitive science as a formal theory of information and computation.

Economics' claim to be a separate science, then, is not a matter of its investigation of a distinctive set of causal regularities, but rather a matter of how Robbins' definition of economics can be combined with Samuelson's revealed preference theory (RPT), minus Robbins' introspectionism. The beauty of RPT for Ross is that it has nothing to do 'descriptively with any real empirical agents' (107). Thus, Robbins' definition of economics as the science that studies human behavior as a relation between ends and scarce means which have alternative uses stands once the term 'human' is dropped (87). This gives the distinctive Robbins–Samuelson argument pattern (RASP).

Ross goes on to agree with me (Davis 2003) that 'if the ontology of mainstream economics is to be defended, somebody has to provide a new concept of *what an economic agent is*' (111), while disagreeing with me that any anthropocentric view of the agent as an individual human being

is possible (157). Indeed, Ross wants to 'eliminate individualism from economic theory proper without at the same time eliminating agents and individual selves from our wider ontology' (111).

Ross' argument is built up from Daniel Dennett's 'multiple-drafts model' (MDM) theory of consciousness. On this view, brains are massive parallel processors, and as Ross puts it are like national economies, for which the key to their stability of response is the coordination of large amounts of distributed information processing without recourse to any centralized executive command function or central site in the brain where everything comes together. 'Aspects of complex problems have to be handled in parallel by distributed teams of subagents, and these teams have to be partially encapsulated from each other with respect to their sharing of information' (235), such that the brain ends up producing multiple drafts of a succession of partial solutions, indeed so as to often 'solve' problems at ineffective levels of abstraction. Further, as there is no self-as-executive, when this on-going process is pushed by the world to report or act, some subset or part of the overall information processing gets privileged ('selected for fame' in Dennett's words), and interpreted as the current content of consciousness. This interpretation or judgment is moreover not a report of the individual's internal states, but involves a 'triangulation among external conditions, proprioperceptive signals, overall behavioral track record, and expectations' (236) regarding whatever occasioned the report or action, and in intentional stance terms is a matter of the subject reflexively taking an intentional stance toward herself.

From this Ross rejects an Aristotelianism that takes human beings as prototypical agents, with Becker's 'anthropocentric neoclassicism' as its best expression in economics (154). Agents are rather defined in Samuelsonian RPT terms as:

> any system that observes certain consistency conditions in behavior, such that it can be interpreted *as if* it is maximizing the value of a function that maps a system of preferences over commodity bundles onto the real numbers. (245)

Ross uses as two defences of this: the money-pump idea and the idea that something must be constant when systems are in continual adjustment to their environment. On the first, 'agents' with cyclical preferences 'disappear from the market *as economic agents*' (246). The latter idea,

> servosystematicity – control of a local entropy through sensitivity to negative feedback – is fundamental to the concept of agency, and has become increasingly emphasized across the cognitive sciences. (248)

I return to these two rationales in my evaluation below.

Behavior of simple biological individuals that can be individuated genetically, such as insects, ideally fits the description of an economic

agent for Ross because of their hard-wired character. In contrast, social animals need to solve highly complicated coordination games that are evolutionary in nature, and as they are drawn into such games they are forced to become more and more sophisticated economic agents. Thus complex sociality is negatively correlated with Samuelsonian economic agency, as manifest in such phenomena as preference reversals and time inconsistency. Indeed, the very idea of a person cannot be Robinson Crusoe because it presupposes this social embeddedness (313). As people become increasingly complex and social, and can less and less be represented in terms of Samuelsonian economic agency, they make increasing use of evolved public signaling systems – 'external scaffolding' (286) – of which public language is the most important form.

The agents in economics, then, are human beings' subpersonal aspects that can be said to behave according to RASP. Ross treats as paradigmatic Glimcher's (2003) neuroeconomics account of individuals made up of subpersonal agents in terms of brain modules and Ainslie's (2001) picoeconomics account of individuals' divergent subpersonal short-range and long-range interests engaged in repeated prisoner's dilemma bargaining games, and integrated with one another through processes like political log-rolling among shifting coalitions. Emphasizing Ainslie, Ross states his own 'book's central thesis' to be 'what a person is: a set of basically compatible long-range interests that have co-opted a sufficient army of short-range interests into their coalition to maintain stable equilibrium' (351). In Ross' game theoretic understanding of this, human beings' subpersonal agents play evolutionary games with one another.

Human beings, then, are communities of these subpersonal RASP agents (like nations in traditional neoclassical analysis) or 'complex assemblies of servosystematic architectures' (256). Dennett's MDM theory of consciousness and intentional-stance functionalism enters into this picture when a human being with the no self-as-executive happens to privilege some subpersonal agent or coalition of agents 'within' herself when in (evolutionary game theoretic) interaction with other human beings similarly privileging some subpersonal agent or coalition of agents 'within' themselves. This brings about adjustment in each person's subpersonal agent coalitions and privileging of particular subpersonal agents (or coalitions of them), so that evolutionary games between people interact with evolutionary games 'within' people.

In contrast to standard thinking, then, the games subpersonal agents play are the subject of microeconomics (the subject of this book), whereas the games whole human beings play with one another are the subject of macroeconomics (the subject of Ross' forthcoming sequel). But as individuals are changed in their interaction with one another, changing their internal interaction of subpersonal selves, and as their subpersonal interaction also changes their interaction with one another, neither the

interior and exterior boundaries of the person, nor the boundaries between macroeconomics and microeconomics are stable.

Finally, the way we see human beings is as continually engaged in constructing their selves by reflexively taking the intentional stance toward themselves. For Ross, this recalls Taylor (1989) and Williams (1976) who see individuals as constantly writing their own biographies as coherent narratives, though Ross emphasizes that because people do this in interaction with others, as authors of their own selves people have multiple co-authors.

RESCUING THE INDIVIDUAL?

I agree with Ross that the unity of the individual needs to be accounted for, is not accounted for in neoclassical economics, and that accounting for the unity of individual needs to be understood in terms of the social interaction of individuals or social embeddedness. Where I disagree with him is in regard to how individuals' subpersonal selves are to be understood and then related to the individual human being as a whole.

Ross's argument is that human beings as complexes of subpersonal aspects are not themselves agents, as evidenced by such things as preference reversals and time inconsistency, but their subpersonal aspects are agents because they are Samuelsonian rational maximizers, who by definition cannot exhibit such things as preference reversals and time inconsistency. Assuming for the moment, that the subpersonal aspects of people are agents, why should they be thought of as Samuelsonian agents?

In his defence of this claim, Ross gives two criteria for being an agent: not being a money pump, or the idea of not being the source of one's own destruction, and servosystematicy, or the idea of a core of stability in the face of adjustment to one's environment. The basis for the first is the intuition that the idea of an agent money pump is self-contradictory. Logically this makes sense, but it also makes sense to imagine there exist agents guided by principles that ultimately insure their destruction, but only after some perhaps lengthy period during which they exhibit activity and plausibly act as agents. So agents need not necessarily be Samuelsonian agents. The intuition behind the servosystematicity idea is that unstable response to the environment changes what the agent is, and thus makes it difficult to re-identify the agent across change. Yes, but that Samuelsonian agents have stable preferences does not reflect servosystematicity because there is no 'control of a local entropy through sensitivity to negative feedback' (246) in such an agent. The agent is simply abstracted from the environment, as Ross makes clear in emphasizing the logical, non-agent specific character of RPT. So neither need agents be Samuelsonian, nor do Samuelsonian agents qualify as agents in the servosystematic sense.

Separately, Ross accepts my argument that vary Becker fails to show that individuals can be understood in terms of utility functions (157, 187) as well as my location of this failure in a replication of Locke's circular argument for personal identity (225). But then how are subpersonal Samuelsonian agents to be identified in terms of their utility functions?

Let us, however, back up the truck, and ask a question that precedes asking why agents should be seen as Samuelsonian, namely, why should the subpersonal aspects of people be thought to be agents at all? It begs the question to say that the subpersonal aspects of people are agents because they have utility functions, because that is Ross' definition of an agent. Ross needs to be able to say these subpersonal aspects are agents, and *then* assign them utility functions. His best argument, it seems, would be to claim that the subpersonal aspects of people are servosystematic. But while he makes a case for saying that insects and other hard-wired simple creatures might be genetically individuated as servosystematic systems, the case does not seem to have been made for the subpersonal aspects of people. Nor is it clear that the two paradigmatic figures Ross discusses in this regard, Glimcher (2003) and Ainslie (2001), have made the case in terms of brain modules and short-range and long-range interests respectively. Neuroscience extensively discusses transference of functions across areas of the brain, often accompanied by modification of function. How does one individuate a brain module then? Short-range and long-range interests might be individuated in general by time frame, but does this take us very far in terms of individuating particular short-range and long-range interests?

If Ross does not make the case for the subpersonal aspects of people being agents, then does the dreaded Aristotelian view that the human being is the prototypical agent get rehabilitated by default? Ross inadvertently opens the door to such a view by allowing that human beings are like countries, which if lacking a clear principle of individuation, are nonetheless taken as some sort of units. Indeed, on a more charitable view of money pump agent idea and with creative application of the servosystematicity idea, countries might well be agents. Then why not human individuals, who are like countries? If the case is not made that agents must be Samuelsonian agents, then the case is not made that agents cannot exhibit preference reversals and time inconsistency, which Ross believes cognitive science has successfully shown to be the case for human individuals.

I agree with Ross, however, that much remains to be done to explain what the unity of the human individual consists in, or indeed the extent to which the human individual can be thought a unity. At the same time, the way forward in this regard might not involve the strategy Ross recommends. He seems to me to be right to argue that seeing interaction between people as evolutionary games impacts on the interaction 'within' individuals between their different subpersonal aspects, and vice versa – and that the boundary between microeconomics and macroeconomics, as he understands these domains, is porous. But if social interaction impacts

the evolutionary games within individuals, why doesn't this also impact what the different subpersonal aspects of the person are? Why, that is, are there atoms or atomistic individuals within the space of the person, just as there are atoms or atomistic individuals in the space of the economy in neoclassicism? Or putting this the other way around, if Ross believes social interaction precludes the standard neoclassical idea that whole persons have stable preferences, shouldn't he also say that interaction between a person's different subpersonal aspects precludes there being stable preferences for subpersonal 'agents'?

I previously suggested a different strategy for a way forward that targets neoclassicism's historic commitment to subjectivism (Davis 2003) in a way recalling Ross' critique of internalism. That strategy, it seems, goes further toward erasing the boundary between what Ross understands as microeconomics and macroeconomics by taking the subpersonal aspects of the individual not as specific to the individual but as inescapably social in the form of the individuals' multiple social affiliations, social identities, social group locations, and social positions. The argument, then, is that the unity of the individual needs to be understood as a capability that might obtain for the individual for transiting across these social subpersonal aspects. But these social subpersonal aspects are not atoms nor simply other individual agents, as in neoclassical market theory, but only different social settings. This approach does not seem too distantly removed from Ross's approach as it might seem. Thus, despite the differences discussed here, I am sympathetic to what I take to be the goals of Ross' argument. And in any event, I strongly recommend his book for its serious engagement with economics' encounter with cognitive science.

John B. Davis
University of Amsterdam and Marquette University

REFERENCES

Ainslie, G. 2001. *Picoeconomics*. Cambridge: Cambridge University Press.

Bowles, S. 2003. *Microeconomics: Behavior, Institutions, and Evolution*. Princeton, NJ: Princeton University Press.

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

Dupré, J. 2001. *Human Nature and the Limits of Science*. Oxford University Press.

Glimcher, P.W. 2003. *Decisions, Uncertainty, and the Brain*. MIT Press.

Hands, W. 2005. Introspection, revealed preference and the explanatory potential of neoclassical economics: a critical response to Don Ross on the Robbins–Samuelson Argument Pattern (RASP) or how I learned to stop worrying and love rational choice theory. Unpublished.

Hausman, D. 1992. *The Inexact and Separate Science of Economics*. Cambridge University Press.

Mäki, U. 1992. Friedman and realism. *Research in the History of Economic Thought and Methodology* 4:127–43.

Nagel, T. 1986. *The View from Nowhere*. Oxford University Press.

Searle, J. 1997. *The Construction of Social Reality*. Free Press.

Shannon, C. 1948. The mathematical theory of communication. *Bell System Technical Journal* 27:37–423, 623–56.
Taylor 1989. *The Sources of the Self.* Cambridge University Press.
Williams, B. 1976. Persons, character, and morality. In *The Identities of Persons*, ed. A. Rorty, 184–205. University of California Press.

doi:10.1017/S0266267107001411

William Stanley Jevons and the Making of Modern Economics, by Harro Maas. Cambridge University Press, 2005, xxii+330 pages.

Harro Maas's book on Jevons is an extremely impressive piece of scholarship – one more than deserving of the Joseph J. Spengler Book Award it recently received from the History of Economics Society. It is well researched, engagingly written, and overall very persuasive.

The book is not a biography of Jevons in the traditional, birth to grave, sense. Maas draws heavily on Jevons's intellectual and social context, but it would not be fair to call the book science studies, since it has neither the site-specific focus of micro-constructivist studies nor the interest-based explanatory strategy of more macro-sociological studies. If it must be labeled, I would call it historical epistemology – an effort to understand how and why Jevons came to consider certain theoretical propositions to be knowledge as a result of his particular personal experiences and general intellectual context.

The central thesis is that Jevons's approach to economic theory – both in his landmark *Theory of Political Economy* (TPE) and in other, more applied, research on various economic subjects – was based on at least four, fundamentally intertwined, commitments. First, a nineteenth-century British notion of mathematics: an applied-scientific notion that tied mathematics inexorably to practical, particularly physics-based, problems. Second, a statistical conception of scientific explananda – where the phenomena to be explained, and the scientific laws that provide the explanations, concern averages and not specific individual events or observations. Third, a commitment to mechanical analogy as an adequate, perhaps the only adequate, scientific mode of understanding. To build a mechanical model, or to capture the essential characteristics of some phenomenon in such a model, was, for Jevons, sufficient for rational intelligibility. Finally, and the point that seems to receive the most attention from the author, Jevons's belief in the substantive identity of the sciences of mind (moral science) and the sciences of matter (physical science). For Jevons, human consciousness was subject to the same type of scientific inquiry that characterized the physical sciences: 'There was no longer any *categorical* distinction . . . between mind and machines' (p. 138). This view distinguished Jevons from those like John Stuart Mill who endorsed a science of mind, but maintained that such a science would