tastes have figured prominently in environmental activism: the farmers' market movement, the "fair trade" imprimatur on consumer goods, the recent efforts to persuade consumers to "buy local," and the rejection of bottled water are all examples. Leopold would certainly not be satisfied with the current scene, but he might be pleased with its direction.

Aldo Leopold was the quintessential multi-disciplinary man, and the new discipline he helped to create, ecology, was made up from the natural, physical, and social sciences. But he paid the price sometimes for this multi-disciplinarity and for venturing fearlessly into foreign disciplinary parts. He ruminated as follows towards the end of his life. "A professor may pluck the strings of his own instrument, but never that of another, and if he listens for music he must never admit it to his fellows or to his students. For all are restrained by an ironbound taboo which decrees that the construction of instruments is the domain of science, while the detection of harmony is the domain of poets" (Leopold 1968, 153). Leopold was both scientist and poet.

## REFERENCES


Barber, William, ed. 1988. *Breaking the Academic Mould: Economists and American Higher Learning in the Nineteenth Century.* Middletown, CT: Wesleyan University Press.

Bentham, Jeremy. 1948. *A Fragment on Government and an Introduction to the Principles of Morals and Legislation*, edited by Wilfrid Harrison. Oxford: Basil Blackwell.

Bentham, Jeremy. 1962. *The Works of Jeremy Bentham*, edited by J. Bowring. New York: Russell and Russell.

Callicott, J. Baird, ed. 1987. *Companion to a Sand County Almanac.* Madison, WI: University of Wisconsin Press.

Club of Rome. 1972. *The Limits to Growth*, edited by. D. Meadows *et al.* New York: Universe.

Costanza, Robert *et al . 1997. An Introduction to Ecological Economics.* Boca Raton, FL: St. Lucie Press.

Ely, Richard T. and George S. Wehrwein. 1940. *Land Economics.* New York: Macmillan.

Galbraith, John Kenneth. 1958. *The Affluent Society.* Boston, MA: Houghton Mifflin.

Hadley, Arthur Twining. 1896. *Economics: An Account of the Relations between Private Property and Public Welfare.* New York: Putnam's Sons, 1897.

Leopold, Aldo. 1933. *Game Management.* Madison, WI: University of Wisconsin Press, 1986.

Leopold, Aldo. 1968. *A Sand County Almanac and Sketches Here and There.* New York: Oxford University Press.

Leopold, Aldo. 1991. *The River of the Mother of God and Other Essays*, edited by Susan L. Flader and J. Baird Callicott. Madison, WI: University of Wisconsin Press.

Leopold, Aldo. 1999. *For the Health of the Land*, edited by J. Baird Callicott and Eric T. Freyfogle. Washington, D.C.: Island Press.

Meine, Curt. 1988. *Aldo Leopold: His Life and Work.* Madison, WI: University of Wisconsin Press.

Newton, Julianne Lutz. 2006. *Aldo Leopold's Odyssey.* Washington, D.C.: Island Press.

Palmini, D. J. 1993. "The Conservation Economics of Aldo Leopold." *Wisconsin Academy Review* 39 (3):37–44.

Vaughn, Gerald F. 1999. "The Land Economics of Aldo Leopold." *Land Economics* 75 (February): 156–59.

Veblen, Thorstein. 1899. *The Theory of the Leisure Class.* New York: Modern Library, 1931.


*Journal of the History of Economic Thought,*
*Volume 30, Number 4, December 2008*

# INTROSPECTION, REVEALED PREFERENCE, AND NEOCLASSICAL ECONOMICS: A CRITICAL RESPONSE TO DON ROSS ON THE ROBBINS-SAMUELSON ARGUMENT PATTERN

BY

D. WADE HANDS

## INTRODUCTION

Don Ross' *Economic Theory and Cognitive Science* (2005) is a very important contribution to the philosophical debate over the nature, significance, and scientific adequacy of neoclassical economics: particularly the relationship between neoclassical theory and contemporary cognitive science. Although I am skeptically intrigued by Ross' general argument, this particular paper will make no attempt to evaluate the success or failure of his overall project. I would also note, a fact that may not be obvious from what follows, that there is much in the book that I agree with, as well as many parts that have substantially changed the way that I think about various issues. It is an important contribution to the literature—much to learn and much to challenge—that will be discussed for a long time. I will only attempt to slice off a small portion for examination here.[1]

The only portion of the argument considered here is the historical part: Ross' defense of the so-called Lionel Robbins and Paul Samuelson argument pattern (RASP). Although he offers various characterizations of RASP in the text, the bottom line is that he considers it to be textbook microeconomic theory—primarily "neoclassical consumer theory" (p. 29).[2] The relevant agents are assumed to have

University of Puget Sound, Tacoma, Washington 98422. This research was supported in part by National Science Foundation Grant 0422823. Helpful comments on earlier drafts were received from Bruce Caldwell, John Davis, Daniel Hausman, Harro Maas, Philip Mirowski, Don Ross, Robert Sugden, and two anonymous referees. I am of course solely responsible for the final result.

[1]I would also note that it is written in a very entertaining style: essentially a debate with three recent contributors to the literature—Davis (2003), Dupré (1993), and Mirowski (2002)—with Mirowski receiving the most attention.

[2]Page numbers without additional reference information refer to Ross (2005).

ISSN 1053-8372 print; ISSN 1469-9656 online/07/04000453-478 
doi:10.1017/S1053837208000436

well-ordered preferences (a utility function) and act optimally (maximize) given the constraints they face (scarcity of means):

> A loose "Robbins-Samuelson" conception like this is, plausibly, thought to be right by many contemporary economists, whatever confused buzz of noises they make when trying to take recent behavioral-experimental evidence seriously. It captures, after all, what they still generally say in undergraduate textbooks when they're trying to simplify matters for students (p. 117).

I will summarize Ross' position in the next section, but the bottom line is that the neoclassical economic agent is not dead; or at least should not be dead according to Ross, because such agents still have a lot to offer behavioral science. Although this is not an unusual position in economics, it is an unusual position for a philosopher, and the *particular way* that he defends this position puts him at odds with almost everyone writing on the philosophy of neoclassical theory. On one hand, there are many critics who use the negative empirical results from recent experimental economics, behavioral economics, and experimental psychology to argue for the elimination of the utility maximizing agent from economics and the other disciplines where it now appears. Ross is certainly not in this camp; he defends both the practical usefulness and scientific adequacy of neoclassical choice theory. He argues that a research strategy based on evolutionary game theory, when paired with the insights from cognitive science that he provides, "preserves rather than displaces the core philosophical assumptions of neoclassicism" (p. 70). On the other hand, there are many defenders of the status quo who insist that none of these recent criticisms represent a serious challenge to the old microeconomic workhorse. Ross is not in this camp either, in part because he finds many of the criticisms raised by experimental and behavioral economics—specifically preference reversals and time inconsistency (pp. 177–89)—to be right on target, and in part because he disagrees with the particular way some economists (specifically Gary Becker) characterize the theory they are defending. The bottom line for Ross, unlike most other commentators, is that neoclassical theory is *appropriate and scientifically legitimate*, even though many of these recent criticisms are *essentially correct*. He argues that in recent years economists have greatly improved the way they model exchange and other social interactions, but, unlike many of those who would agree with this view of recent developments, he also believes that utility maximization is just as useful in this new theoretical environment as it was in the Walrasian milieu of the mid-twentieth century.

As noted above, I will not attempt to challenge (or defend) Ross' general position about how well neoclassical theory fits (or can be made to fit) in with recent theoretical developments such as behavioral economics and evolutionary game theory. Nor will I challenge (or defend) the various philosophical resources that Ross enlists in his defense of rational choice theory, such as Daniel Dennett's "intentional-stance functionalism." My subject will be Ross' characterization of Robbins and Samuelson—in particular, his characterization of how these two economists justified (or how they should/could have justified) the particular versions of rational/consumer choice theory they supported. My argument will be that whether one accepts or rejects Ross' claims about neoclassical theory, its relationship to contemporary cognitive science, or Dennett's intentional stance, his extensive discussion of Robbins and Samuelson should be independent of that assessment. These are simply not resources that can be used persuasively in this way if one maintains fidelity to the arguments made by Robbins and/or Samuelson. Ross is not defending just any-old version of neoclassical theory—he is defending *a particular version* (RASP) that will be able to do certain things, accommodate certain recent theoretical developments, live up to certain philosophical standards, allow for a broad characterization of agency, and so forth—and my argument is that Ross' version of rational choice is simply not in the writings of either Robbins or Samuelson (and is, in many respects, in conflict with both the economic theories they advocated and the philosophical defense they provided for those theories).



## I. THE FEATURES OF A SCIENTIFIC ECONOMICS WORTH HAVING

In this section I will discuss a few of the features of neoclassical choice theory (NCT)[3] that are key to Ross' defense of it as (part of) a sound predictive and explanatory strategy within contemporary economics (RASP). I will only discuss three of the many things that Ross considers to be prerequisites for a scientifically successful NCT, and my selection criterion is the relevance of these particular features to Ross' claims about the Robbins-Samuelson heritage of the research program.

### *It Is Systematic*

It seems obvious that any reasonable scientific theory or explanatory strategy should be systematic in the sense of being step-by-step or methodical, and although there are passages where Ross employs the term in this general sense, he usually has a more specific definition in mind. For Ross, "systematic" essentially means that the relevant organizing principles are "independent of any particular human purposes" (p. 21) and thus constitute a "view from no particular human place" (p. 23). It is important to note that the view from no particular human place is not the same as the view from nowhere—one does not need to have a God's eye view to be systematic in Ross' sense.

Ross is explicit about his definition of systematicity, but he often uses mathematical, particularly axiomatic, formalization as a proxy for such systematicity. Although he never clarifies the exact relationship between axiomatic structure and systematicity, he does consistently use such formalization as evidence that the theory in question systematic. Although I think a separate paper could be written about Ross' treatment of systematicity—particularly to sort out the relationship between

[3]It is not a coincidence that I have used a different acronym for neoclassical choice theory than Ross uses. Ross uses RPT throughout, for "revealed preference theory." The problem is that for economists "revealed preference theory" identifies the research program associated with the particular axiom (what came to be called the Weak Axiom of Revealed Preference—WARP) first presented in Samuelson (1938). Ross does not use the term in this narrow sense; he uses it to subsume all of neoclassical choice theory from the ordinal revolution, through Samuelson's WARP, and on to Arrow-Debreu general equilibrium theory. As he says: "By 'RPT' I want to refer to the entire package of Samuelsonian theoretical commitments, not just to what is formally necessary for the axiomatization of 'choice as revealed preference'" (p. 105) and "Perfection of generality and elegance was achieved by Debreu (1959); and it was in this axiomatization that RPT became the standard textbook model of the economic agent" (p. 112). Economists often expand the definition of revealed preference to include the "strong axiom" introduced by Houthakker (1950) and the surrounding literature, but not all of neoclassical choice theory (see any standard textbook, Mas-Colell, Whinston, and Green (1995) for example).

"independent of any particular human interests" and "axiomatization"—this is not that paper. For the purposes here I will just accept Ross' notion of systematicity and rely on axiomatization as a reliable proxy for it. The main point is simply that for Ross science needs to be systematic and NCT is.

*The "Agents" In Neo-classical Choice Theory Are Not Restricted To (Or Even Best Instantiated By) Human Beings*

Economists have of course traditionally assumed that the agents appearing in their models represented flesh-and-blood human beings (or ensembles of such humans: firms, unions, governments, etc.), but Ross argues that such anthropocentrism is an inappropriate, and actually quite problematic, way of thinking about the relevant theory. Ross' interpretation of NCT allows for a wide range of possible "agents," and this diversity counts as a big plus for the theory's scientific credentials. The potential set of rational "agents" includes certain computers and the associated software, humans that are normally considered pre-rational such as newborn infants (p. 292), a variety of non-human life-forms such as insects (pp. 251, 331) and pigeons (who behave "like good Samuelsonians" p. 277), and, following Glimcher (2003), neurons within the human brain itself (pp. 320–34). The argument is not simply that NCT "can" be applied to such non-traditional agents, but rather that many of these agents, insects in particular, actually behave in a more prototypically neoclassical way than most flesh-and-blood humans. This of course means that NCT need not be "individualistic" in the way that economists have typically characterized microeconomics as individualistic: requiring all explanations of social/market phenomena be reducible to the rational actions of individual (human) agents.[4] As Ross explains, we need "a form of economic analysis that we can apply to patterns of exchange in any market whatsoever, regardless of whether the agents that comprise it are people, animals, firms, countries, or computers selling derivatives to each other, and regardless of whether they calculate their goals and strategies themselves" (p. 247).

*It Is Consistent With Dennett's Intentional-stance Functionalism*

One much-discussed solution to the various problems associated with intentional (or folk-psychological) explanations of human behavior is Daniel Dennett's intentional-stance functionalism (Dennett 1987, 1991a, and elsewhere). Folk psychology is simply a convenient label for explanations in terms of the beliefs and desires/goals of the relevant agent. For example, how do we normally explain something like "Bob went to the movies" (or more generally that "agent x did action A")? A reasonable explanation is that Bob had a desire for some entertainment, he believed that the best way to fulfill that desire given the various constraints he faced (time, money, etc.) was to go to the movies, and therefore Bob went to the movies. Although such belief-action-desire (BAD) explanations are the stuff of everyday life (and law, and fiction,

[4]Both Davis (2003) and Mirowski (2002) make similar arguments about neoclassical agents not being restricted to human beings and neoclassical economics not being as individualist as commonly supposed. The difference, as Ross discusses, is that for these authors this is a failure of NCT, while for Ross it is a virtue.



and romance), they have myriad well-known difficulties *as scientific explanations* of human behavior.

It is not necessary to review the various problems that have been suggested regarding the epistemic credentials of such folk psychological explanations,[5] but it is useful to discuss one particular, much-debated, solution to these problems. The solution is "eliminative materialism"—the view that beliefs, desires and other mental states simply have no place in serious science and should be totally eliminated and replaced by scientific neuroscience (neurophysiology and brain chemistry).[6] Eliminativism is in many ways a version of behaviorism—they both reject mental state explanations as scientifically respectable—but it goes even further than behaviorism by specifying exactly what form a proper non-BAD explanation of behavior must take. The elimination of beliefs, desires, and other mental states from the list of legitimate scientific entities, effectively eliminates "the mind" (as opposed to the brain) from the list of things that might be causally involved in the scientific explanation of human behavior.

> "Minds" might then continue to be useful points of reference for sloppy, everyday getting along—in the same way that most people continue to talk about the sun rising even when they know it's our horizon that's doing the relative moving—but, in fact, there are no minds. Just as some of our ancestors talked about social properties of inconvenient groups of women by reference to "witches," so we rationalize our behavior through loose fantasies about minds and mental causation. But there aren't really, any witches (and never were), and there aren't, really, any minds or mental causation (p. 191).

Such "eliminativism" is a theme that runs throughout Ross' book. It is not a view that he supports, but it is a view that is often associated with positions that he does support—which puts him in the position of repeatedly needing to explain why his argument does not imply, or require, eliminativism. The reason is of course because NCT can be viewed as a version of folk psychology—a narrow and very mathematically formalized version, but folk psychology nonetheless. The standard rational choice explanation starts with the preferences or utility function of the agent (the agent's desires) and the various constraints the agent faces (the agent's beliefs about such constraints). The behavior is then explained as the result of instrumentally rational action (optimization) given these beliefs and desires. If one wants to defend such rational choice explanations in light of recent developments in cognitive science, then one needs to have a reply to eliminativism, particularly since much of the work in experimental economics and game theory that Ross wants to defend has "an eliminativist ring" (p. 193) to it.

Enter Daniel Dennett: Ross argues that Dennett's intentional-stance functionalism provides a way around these issues. In his words: "*eliminativism* won't work to legitimize an economic science, but Dennett's theory will" (p. 46). In other words, Dennett's intentional-stance functionalism allows Ross to steer NCT safely along a path between eliminativism and intentional folk psychology.

[5]Ross discusses a number of these issues; also see Rosenberg (1992, 1995) for a particularly clear discussion of the various problems with folk-psychological science.
[6]Canonical statements of the position are Patricia Churchland (1986) and Paul Churchland (1984).

In order to understand how Ross applies intentional-stance functionalism, it is useful to review Dennett's notion of the "intentional stance." Dennett often discusses three basic approaches to predicting and explaining the behavior of a system or entity: the physical stance, the design stance, and the intentional stance. When one takes the physical stance they explain the behavior in question in terms of the laws of physical (including biological) science. The design stance explains the behavior as a result of what the system was "designed" to do (in science this usually means "designed by us" but the design stance is also present in theological explanations). Finally, one is taking the intentional stance when they explain the relevant behavior as the result of rational action on the basis of beliefs and desires.

Notice that the intentional stance does not require the agent to "really" have beliefs and desires sloshing about in their head. All that is required is that it provides a way of making reliable empirical predictions about the relevant agent's behavior (more reliable than available from alternative stances). This has led many commentators to interpret (and criticize) Dennett's intentional stance as a purely *instrumental* approach to behavioral science: the view that the adequacy of a scientific theory is based solely on how well it predicts—saves the empirical phenomena—and thus has nothing to do with whether its theoretical terms correctly identify real features of, or causal mechanism within, the world. Since instrumentalism is the standard interpretation of Milton Friedman's famous paper on economic methodology (Friedman 1953), an instrumentalist reading of Dennett's intentional stance would essentially reduce Ross' philosophy of economics to an updated version of Friedman's methodology.

Even though "Dennett has struggled to disown" (Ross 2000a, p. 19) the instrumentalism of his position—and the jury is still out on the question within contemporary philosophy of mind—Ross insists on a realist interpretation of the intentional stance. Dennett denies his approach is instrumentalist and has offered his own explication in terms of "real patterns" (Dennett 1991b). Ross has gone even further in the realist direction, elaborating his own Dennettian view termed "rainforest realism" (Ross 2000b) and that is the ontology he builds into the version of NCT he defends.

The bottom line is that Ross wants to defend NCT as a successful scientific approach to predicting and explaining economic behavior, but he also wants it to accommodate at least two facts that cut hard against it literally being a realistic description of the actual processes that precipitate economic behavior. First, it needs to be applicable to insects, pigeons, individual neurons, computers, and other entities that do not seem to have intentional mental states; and second, he is willing to accept the literature from experimental psychology and experimental economics that offers convincing empirical evidence that actual human agents seldom make choices in the way that NCT claims they should. Intentional-stance functionalism allows him to steer his way through these two potential pitfalls. If all it means to *have* intentional states is "to exhibit behavioral patterns that can't be predicted or explained without recognition of the patterns indexed by the intentional states in question" (p. 63) then the mathematical folk psychology that is NCT may be just fine. It is of course an empirical question whether there exists a better way to predict or explain the behavior of insects, chess software, or humans in the grocery store, but if we are willing to grant him the empirical effectiveness of NCT, the deed is done. Intentional-stance functionalism thus allows him to "talk about *intentional* states—beliefs and desires—as real objects for scientific investigation, without having to suppose that they pick out internal states of individual people" (p. 69). People do not need to really have preferences, utility functions, or even minds, for NCT to legitimately be applied to their behavior. All that is required is that such an intentional-stance provides the best (perhaps only) predictions and explanations of what they do, and yet, if one is willing to grant the Ross Dennettian realist interpretation of the intentional stance, the resulting scientific theory also seems to be safe from the charge of instrumentalism (and thus just being an updated version of Friedman's methodology).



## II. LIONEL ROBBINS AND RATIONAL CHOICE

So how does Lionel Robbins, and particularly his influential *An Essay on the Nature and Significance of Economic Science* (1935), fit into Ross' systematic, dehumanized, Dennettian, view of NCT? Why does he call it "Robbins-Samuelson" NCT? Before answering I will begin by reviewing three well-known aspects of Robbins's position. Then I will turn to Ross' specific discussion of how these three aspects relate to the issues discussed in the previous section and challenge that interpretation.

Robbins's famous definition of economics is found on page 16 of the second edition of his *Essay.*[7] It is the definition that still appears in the first chapter of most introductory textbooks: "Economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses" (Robbins 1935, p. 16).

As Ross points out, this definition is an important shift from the previous British tradition of Mill and Marshall in that economics need not have anything to do with wealth, money, or even markets. It is a scarcity-based definition of the discipline; if there are scarce means for achieving the relevant goals, it is an economic problem. As Robbins says: "It does not attempt to pick out certain *kinds* of behaviour, but focuses attention on a particular *aspect* of behaviour, the form imposed by the influence of scarcity" (Robbins 1935, pp. 16–17). Obviously—not noted by Robbins of course, but obviously nonetheless—insects and other non-humans face scarcity.

In addition to the concept of scarcity, this definition requires the economic agent to have "ends"; it is "scarcity of *given* means for the attainment of *given* ends" (Robbins 1935, p. 46). In consumer choice theory these ends will involve preferences and utility, in the case of a firm it will be profit, but in any case the ends will need to be sufficiently well-ordered that *choice* is possible. Thus for Robbins, economic decisions always involve choice and opportunity cost:

> But when time and the means for achieving ends are limited *and* capable of being distinguished in order of importance, then behaviour necessarily assumes the form of choice. Every act which involves time and scarce means for the achievement of one end involves the relinquishment of their use for the achievement of another. It has an economic aspect (Robbins 1935, p. 14).

[7]There are differences between the first and second editions of Robbins's *Essay,* but since these differences do not matter to the argument at hand I will refer exclusively to the second edition. See Howson (2004) for a detailed discussion of the background.

The second relatively non-controversial aspect of Robbins view of the nature and significance of economics discussed by Ross is Robbins's *introspectivism*. The introspectivism enters in response to the obvious epistemological question raised by Robbins's definition of economics: How do we know there are such economic problems? How do we know there is scarcity and that people have well-ordered preferences? The answer to the scarcity part of the question is straightforward; for Robbins it is just an obvious empirical fact that scarcity exists and he uses examples of non-scarce goods (such as air) to emphasize the point (Robbins 1935, p. 15). The ordering of importance (preferences) that people have in their heads, on the other hand, is quite a different matter. Here our knowledge is not based on observation of others, but rather on self-knowledge, on introspection:

> [T]he foundation of the theory of value is the assumption that the different things that the individual wants to do have a different importance to him, and can be arranged therefore in a certain order. This notion can be expressed in various ways and with varying degrees of precision . . . But in the last analysis it reduces to this, that we can judge whether different possible experiences are of equivalent or greater or less importance to us (Robbins 1935, p. 75).

Unlike the facts of natural science, which require experimentation, the existence of preferences is a matter of everyday (inner) experience:

> The main postulates of the theory of value is the fact that individuals can arrange their preferences in an order, and in fact do so . . . We do not need controlled experiments to establish their validity: they are so much the stuff of our everyday experience that they have only to be stated to be recognised as obvious (Robbins 1935, pp. 78–79).

The final aspect of Robbins's position to emphasize is the most enduring legacy of his work: the argument for the impossibility of making interpersonal utility comparisons or the incommensurability of individual preferences. The general acceptance of Robbins's position on this matter had a substantive impact on the way that most economists viewed the contribution of "scientific" economics to policy debates.

Robbins's argument was that scientifically acceptable postulates can either be grounded in things that are interpersonally observable (like scarcity), or introspectively observable (like preferences), and that interpersonal utility comparisons are neither. They take place inside the head, but it is someone else's head, and as such it "is a comparison which necessarily falls outside the scope of any positive science" (Robbins 1935, p. 139). "Introspection does not enable A to measure what is going on in B's mind, nor B to measure what is going on in A's. There is no way of comparing the satisfactions of different people" (Robbins 1935, p. 140).

When Robbins is discussing the question a few years later (Robbins 1938), he makes it clear that the issue is not simply that utility functions are not interpersonally observable—since he consistently argued that our knowledge that people order their choices does not come from such observations; it comes from introspection—rather, the problem is that they are not discernible from *either* observation or introspection. He says, "The assumptions of the propositions which did not involve interpersonal comparisons of utility were assumptions which had been verified by observation or introspection" (Robbins 1938, p. 637) and "I still cannot believe that it is helpful to speak as if interpersonal comparisons of utility rest upon scientific foundations—that is, upon observation or introspection" (Robbins 1938, p. 640).



Ross has written about Robbins in previous work (Ross 1999), and he clearly identifies all three of these important features of Robbins's position; I have no quibbles here. My criticisms concern how Ross presents Robbins on positivism and systematicity, and those criticisms spill over into his claims about how Robbins's economic theory relates to non-human agents and Dennett's intentional-stance functionalism. I will start with positivism.

Ross clearly wants to present Robbins's position as "positivist"—consistent with the philosophy of logical positivism—and needs to weave a rather elaborate philosophical tapestry to make the case. On the face of it, Robbins's introspectivism and subjectivism—aspects of Robbins's position that are repeatedly, and correctly, emphasized by Ross—seem to be decidedly at odds with positivist philosophy of science. Logical positivism is generally considered to be an *empiricist* philosophy of science, and Robbins's characterization of the introspective "knowledge" of our own mental states seems much closer to Austrian *a priorism* than any version of empiricism. This is not to say, of course, that one cannot base economic knowledge on introspection and still maintain a fundamentally empiricist epistemology—that is precisely what John Stuart Mill did—but one of the changes that took place within empiricist philosophy during the period between Mill and the Vienna Circle was a change in what counted as an acceptable "observation." By early in the twentieth century introspection was no longer an acceptable source of empirical evidence—observations needed to be "objective" or "inter-subjective"—and that seems to be directly at odds with Robbins's position that the properties of preferences are "know to us by immediate acquaintance" (Robbins 1935, p. 105). Given all this, how is it that Robbins becomes a positivist in Ross' story?

The move is accomplished by appealing to one of the recent reinterpretations of positivist philosophy of science. In the third quarter of the twentieth century, when Thomas Kuhn and others were attacking the so-called "Received View" within the philosophy of science, positivism was broadly considered an empiricist philosophical program; a view supported by early English-language interpreters such as A. J. Ayer and Bertrand Russell, popular surveys such as Suppe (1977), and the writings of certain key figures within the movement (such as Carnap's autobiography 1963). Differences existed between the positivism of original Vienna Circle associated with philosophers like Rudolf Carnap, Otto Neurath, and Moritz Schlick, and the later—and Anglicized—"logical empiricism" of Carl Hempel and others, but the standard view was that a fundamentally empiricist conception of scientific knowledge was common to both. This has begun to change in recent years. The work of Nancy Cartwright (Cartwright, Cat, Fleck, and Uebel 1996), Michael Friedman (1999, 2001), Alan Richardson (1998), Thomas Uebel (1992), and others has now made it clear that positivism is much less homogeneous than had previously been presumed. It seems that differences existed both among various members of the Vienna Circle at particular points in time and between the positions defended by particular positivists over time/texts; as a result, the notion of a positivist "Received View" is now much less well-established than a few decades ago. Although the main theme in this recent literature has been variation and diversity, the one argument that has consistently emerged from the literature is that logical positivism was generally more naturalist and less empiricist than had previously been supposed. One of these recent reinterpretations of early positivism is the neo-Kantian reconstruction of Carnap's early work by Friedman (1999, 2001) and Richardson (1998). Ross uses this neo-Kantian interpretation of positivism, particularly Friedman (1999), to classify Robbins's introspectionism as positivist.

On Friedman's reading an important feature of Carnap's work was its dynamic neo-Kantian position: in opposition to both traditional (Humean) empiricism and Kant's (static) synthetic *a priori*. In Kant's view, the conditions for the possibility of knowledge resided in universal (thus static) transcendental structuring concepts/categories within the human mind—particularly those associated with space, time, and causality—but by the beginning of the twentieth century developments in non-Euclidean geometry and Einstein's general theory of relativity had significantly undermined the universality of these (Euclidean and Newtonian) concepts/categories. The neo-Kantian movement was an attempt to preserve the fundamental Kantian notion of structuring concepts/categories while accommodating these recent scientific developments. The basic idea was to dynamicize these structuring concepts so they would still be constitutive of our knowledge but no longer frozen in Newtonian amber. The result would still be a Kantian (and not Humean) view of scientific knowledge, but one that was less universally transcendental, more flexible, and more consistent with the best scientific practice (and thus more naturalist).

Friedman's argument is that Carnap (particularly in the *Aufbau*) put a formal-logical spin on this neo-Kantian program. The responsibility for objectivity and the possibility of knowledge no longer resided in the particular (Newtonian) content of our mental categories as it had with Kant, but rather in the *form*—the logical structure—of the formal relations involved in our scientific theories. It was, to use Richardson's expression, a "structuralist account of objectivity" (Richardson 1998, p. 29), where the logical-mathematical formalism does all the epistemic work. As Friedman explains:

> The primary problem is to account for the objectivity of scientific knowledge, and the method of solution is based on a form/content distinction. *Scientific knowledge is objective solely in virtue of its formal or structural properties*, and these properties are expressed through the "places" of items of knowledge within a single unified system of knowledge. The project is not *strictly* Kantian, because the notion of form or structure in question here is a purely logical one, understood solely in terms of formal logic. (Friedman 1999, pp. 98–99, emphasis added).

The key to the objectivity of science is thus neither in its grounding in sense data (empiricism) nor in Kant's (Newtonian) synthetic *a priori*, but rather in the structure imposed by mathematical logic. And this—*as a neo-Kantian early positivist—is how Ross interprets Robbins*. Robbins's ideas about the scope and testability of economics are based on "a set of positivist theses about the epistemology of science" (Ross 2005, p. 88) and his view of economics and psychology is "firmly in the heartland of early (Kantian) positivism" (Ross 2005, p. 89). "Robbins articulates a conception of economics according to which it *exactly fits* the early positivist ideal for science. That is, economics cooked to Robbins's recipe is a self-contained deductive structure resting on an introspective foundation" (Ross 2005, p. 90, emphasis added).

Ross also insists that Robbins supported the formal axiomatization that is key to the structuralist neo-Kantian interpretation of positivism: "Positivists typically interpreted



this as requiring that a theory, to be scientific, had to be capable of expression as a formally axiomatized system of generalizations. Nowhere in the *Essay* does Robbins explicitly go quite this far. However, he often hints that a full axiomatization and universality of generalizations is an appropriate limiting idea" (pp. 88–89).

I have two main criticisms of this interpretation of Robbins. The first concerns Robbins's systematicity—the role of formal/logical structures in Robbins's characterization of economics—and second involves the possibility of de-anthropomorphizing Robbins's economics. I will consider systematicity first then turn to non-human choice.

The problem with trying to connect Robbins to positivism through neo-Kantianism is that neither the argument for the necessity of axiomatization, nor the argument that logical structure represents the key to scientific objectivity, are in Robbins's work. As the above quote from Friedman make clear, the formal structure of science does all the philosophical heavy-lifting in the neo-Kantian version of positivism. Robbins never emphasized the mathematical structure of economics—he was less concerned about formalism than many economists of his day—and he certainly never believed that the discipline's scientific validity rested solely on its logico-mathematical structure. There are only a couple equations in Robbins (1935) and the tone of the argument isn't even the verbal-but-axiomatic-in-spirit of some of his predecessors.[8] Economics is valid for Robbins because it starts with two facts: the fact of scarcity which we learn from the world by observation and the fact that people have preference orders which we learn from introspection. Sure, it might be useful to use a little calculus once in a while to make an economic argument, but the discipline's scientific legitimacy certainly does not reside in such mathematical manipulations. There is an ongoing debate about the degree to which Robbins was influenced by British (empiricist) positivism, Smith and the Scottish Enlightenment, Mises's Austrian *a priorism*, and a number of other sets of ideas, but there is no case to be made for Robbins as an advocate of formal-structure-as-objectivity or Bourbakian Arrow-Debreu axiomatization in economic theory.[9]

Regarding the question of non-human agency, I want to argue that Robbins's position cannot accommodate non-humans and is inconsistent with intentional-stance functionalism. The main problem, and it is a problem that Ross conveniently evades in his discussion of Robbins's introspectivism, is the problem of *other minds*. Even if we completely accept Robbins's argument that we know that *we* have ordered preferences, we do not know, given the impossibility of interpersonal knowledge, that *others*—and that includes everyone whose behavior economists are trying to predict and explain—have preferences, or even minds. This is a problem stemming from

[8]It is possible to make the case that Robbins's definition facilitated the later formalization of economics (Backhouse and Medema 2009), but this is a quite different from the claim that Robbins believed that scientific objectivity was solely a matter of logical form.

[9]A case *could* be made to connect the neo-Kantian version of positivist philosophy to the structuralist philosophy of science of Joseph Sneed (1971), Wolfgang Stegmüller (1976), and others; this philosophical framework has been used to reconstruct the structure and dynamics of Arrow-Debreu general equilibrium theory (for example, Stegmüller, Balzer, and Sophn (1982), Balzer and Hamminga (1989)). I discussed this literature in Hands (1985) and pp. 341–49 of Hands (2001). One could thus go from neo-Kantian positivism to *a version* of neoclassical economics, but it is the most formalized version of neoclassicism and has nothing to do with the kind of economic theory that Robbins either endorsed or produced.

Robbins's double commitment to introspectivism and his view that one cannot know the contents of other minds.

On the face of it, it seems very strange that anyone would try to fit Robbins's position into Dennettian framework. Dennett was a student of Gilbert Ryle and insists that his work is broadly in the tradition of his mentor. One can debate the fine points of Ryle's *The Concept of Mind* (1949), but the one most obvious thing to take away from the book is that the whole "ghost in the machine" story—a story grounded fundamentally on introspection and the occult character of other minds—is, to use one of Ross' great phrases from another context, "zanily implausible" (p. 264). And the ghost in the machine story *is Robbins's story*. Two of the key features of Robbins's position are that *introspection* is a unique source of knowledge—as Ryle put it: that "Sense-perceptions can, but consciousness and introspection cannot, be mistaken or confused" (Ryle 1949, p. 14)—and that the contents of other minds (if they exist) are not knowable and are thus not interpersonally comparable—again in Ryle's terms that "one person has no direct access of any sort to the events of the inner life of another" (1949, p. 14).

Robbins not only argues that we have privileged access to our own mental states (preferences) and that we do not have direct, or *quantitative*, access to the preferences of others, he also argues that we *do in fact have access to the qualitative features* of other minds. We do not know the exact preferences of those around us, but we do know that they have preferences and that those preferences have certain qualitative properties (completeness, transitivity, etc.). Introspection is the only reliable source of information about preferences—that is how we know that *we* have them—and we have no access to the minds of others, and yet we know not only that they have preferences, but that they are well-ordered. If we maintain strict compliance with the ghost in the machine and the privilege of introspective knowledge, then (even if somehow we accept there are other minds), we should know nothing about them. In Ryle's words: "If the doctrine of the ghost in the machine were true, not only would people be absolute mysteries to one another, they would also be absolutely intractable" (1949, p. 114). Robbins of course does not say that others are "absolute mysteries" or "absolutely intractable," but to avoid doing so requires him to leap out of the epistemic cage he has built for himself/us.

Since one of Robbins's key assumptions is that *others* have ordered preferences—without this there is no economics as he defines it—he certainly needs to explain how this could possibly be the case given his ghost in the machine commitments. Robbins's implicit answer is that *we know that others have ordered preferences because they are like us*. We know something about humans because we are humans, and we know from introspection that humans have ordered preferences, and therefore it is safe to build the science of economics on the "fact" that all humans (at least the non-infant, non-mentally ill, not-in-a-coma, humans like us) have ordered preferences. This of course makes it an impossible theory to apply to entities that are not in this sense like "us"—actually that is not true, one can apply it instrumentally if it predicts accurately, but if you are not an instrumentalist (as Ross insists) and you want to do what Ross is trying to do—rationalize a version of NCT that is systematic, works effectively for non-humans, and is consistent with intentional stance functionalism—then Robbins doesn't provide any help.

Rationalizing Robbins's various commitments requires some questionable philosophical baggage, but even if one were willing to accept these commitments, one still does not end up with a rational choice theory that is either consistent with Dennett's philosophy of mind or suitable for infants and insects. Now I realize that much of what has been said sounds critical of Robbins—at least Robbins as a philosopher of mind—but that is not my intent. Understanding Robbins's success is a very interesting question in historical epistemology, but it is not the task here. As Ross says, he is not interested in trying to understand history: "Who, aside from an intellectual historian with a librarian's passion for sorting thinkers into constructed boxes, cares?" (p. 114). What he is trying to do though, is to build a modern, cognitive-science-inspired version of rational choice theory for the twenty-first century, and my point is simply that such a construction project will not be successful if Lionel Robbins's defense of choice theory is used as one of the foundational pillars. One could, of course, just pour the modern philosophical concrete and then after the fact put a sign on one of the pillars that says "Robbins"—which often seems, inexplicably, to be what Ross is doing—and it would probably be harmless. But if one really intends the foundation to be constructed in a way that is consistent with the majority of the words that Robbins wrote, then choosing this particular pillar does not seem to be a good idea. With this said, it is time to turn to the second of Ross' pillars, the revealed preference theory of Paul Samuelson.



## III. SAMUELSON, REVEALED PREFERENCE, AND RATIONAL CHOICE

Before embarking on the discussion of Samuelson's revealed preference theory, let me just note two things that make the approach in this section a bit different from the approach in the section on Robbins. First, this section will rely heavily on the existing historical literature on Samuelson, including: Hands (2006), Hausman (2000), Lewin (1996), Rosenberg (1992), and most importantly, Wong (2006). The second point is that my general approach to the critical discussion will be quite different. In my discussion of Robbins, I went through the list of three desirable features—systematic, non-anthropomorphic, and intentional-stance functionalism—and examined problems associated with each (and various combinations). In the Samuelson case, I will approach the subject more holistically, by going directly to the final chapter of Ross' book where he explains in detail how the RASP is supposed to work in the scientific study of economic behavior. All three issues are still involved, but they are homogenized relative to the previous discussion. The bottom line will, of course, be similar: even if one were to accept Ross' arguments about how to engage in RASP-based inquiry, and his claims about the epistemic benefits of doing so, it still has nothing to do with Samuelson or revealed preference.

Ross' discussion of Samuelson relies on two parts of Samuelson's massive professional output: his revealed preference version of consumer choice theory originally presented in Samuelson (1938) and his influential *Foundations of Economic Analysis* (1947). In Ross' story no distinction is made between revealed preference theory and *Foundations*; for Ross these are not "two parts" of Samuelson's research, but rather one relatively homogeneous framework: the approach to neoclassical economics that he calls "revealed preference theory" (RPT). I will argue that the revealed preference project is in fact different than the *Foundations* project and that the differences matter for the question of whether Samuelson's economics has in place in the economic research program that Ross supports (RASP). Not only can revealed preference theory be separated from *Foundations*, neither—revealed preference nor *Foundations*—fits Ross' vision for economic theory.

There are four parts to my discussion of Samuelson. I begin with the conventional wisdom about revealed preference theory. Then I turn to the discussion of RASP where Ross provides a convenient "how to" guide for what he considers proper economic practice. Third, I will return to Samuelson's revealed preference theory and discuss some issues that go beyond the standard story and complete the argument that Ross' vision of NCT is entirely independent, and frequently at odds with, Samuelson's theory of revealed preference. In the final section I will turn to *Foundations* and demonstrate that it is inconsistent (and actually conflicts with) the type of economics that Ross calls RASP.

The standard story about Samuelson's revealed preference theory is roughly as follows. Samuelson's original paper on revealed preference theory (Samuelson 1938) was an attempt to drop even "the last vestiges of the utility analysis" (Samuelson, 1938, p. 62). It was an effort to move consumer choice theory completely away from the notion that consumers have preferences or utility functions governing their behavior, and to replace the entire earlier theoretical apparatus with a purely *behaviorist* theory. The consumer's decision-making process was effectively black-boxed, taking with it the troublesome mental states associated with BAD rational choice explanations. We simply observe the consumer purchasing certain goods at certain prices, and different goods at different prices; if those choices are consistent with the so-called weak axiom of revealed preference (WARP)[10] then Samuelson (1938) demonstrated that all except one of the standard Slutsky restrictions hold on the corresponding demand functions. In other words, it is not necessary to assume that consumers are maximizing (even ordinal) utility; if their observed behavior is consistent with WARP then the main results of ordinal utility theory hold without any reference to utility or preference, and consumer choice theory can be reconstructed on an observational basis that was entirely free from introspection or other scientifically questionable concepts. The late 1930s were the heady days of operationalism and behaviorism, and this solution seemed to be just the gulp of scientific fresh air the profession needed.

The next round of developments in revealed preference theory involved the development of the so-called strong axiom of revealed preference (SARP) by Hendrik Houthakker (1950). Samuelson's original paper demonstrated that WARP (combined with the other assumptions of the model) implied two of the three standard results from ordinal utility theory: the negative semi-definiteness of the Slutsky matrix $[x^T Sx \leq 0 \, \forall x \neq 0]$ and nonpositive own Slutsky substitution terms $[S_{ii} \leq 0 \, \forall i]$, where $[S_{ij}]$ is the nxn Slutsky matrix with representative term:

$$S_{ij} = \frac{\partial x_i}{\partial p_j} - x_j \frac{\partial x_i}{\partial M}.$$

The stronger SARP condition added the third result from the standard theory, the symmetry of the Slutsky terms $[S_{ij} = S_{ji} \, \forall i \neq j]$. A variety of additional results were obtained in the revealed preference literature that followed (for example, Kihlstrom, Mas-Colell, and Sonnenschein 1976 and Uzawa 1971) but they involved weakening the mathematical restrictions of the earlier results rather than providing additional theoretical implications. Since the stronger versions of revealed preference theory have implications identical to the standard ordinal utility theory, the contemporary literature presents the two approaches as alternative ways of achieving the same basic results in consumer choice theory (see for example Mas-Colell, Whinston, and Green 1995, pp. 3–6).

Ross' discussion of revealed preference in chapters three and four is for the most part consistent with this standard story. Recall that Ross labeled Robbins an early neo-Kantian positivist—a claim that I questioned—but he labels Samuelson a behaviorist and a "generic late positivist" (p. 111) and that label fits quite well (at least for Samuelson 1938). Ross also argues that although Samuelson never used the contemporary language of "eliminativism," his approach points naturally in that direction (p. 110, p. 135); this also seems to be accurate, particularly, as I will argue below, in his first revealed preference paper where his purpose was to eliminate all reference to preferences or other BAD notions. Overall, Ross characterizes Samuelson as someone who tried to move choice theory in the behaviorist direction; in consumer choice theory this took the form of substituting the notion of (observable) choice for (unobservable) preference, and this view is broadly consistent with the standard reading of revealed preference theory. My main criticisms do not arise here.

In order to understand my criticism of Ross' interpretation of Samuelson it is useful to turn to the section where he explains how it is, exactly, that the Robbins-Samuelson Argument Pattern is supposed to work—how RASP is used to predict and explain. Applied-RASP is defended as exemplary social science: in part because it epitomizes the three desired characteristics (systematicity, non-human agents, and intentional-stance functionalism), in part because it is consistent with recent developments in cognitive science, and in part because it is just good science. Recall that I am not arguing that Ross' RASP version of NCT is inadequate or that it is not consistent with the three desirable properties he attributes to it, but rather that whatever its epistemic standing or particular characteristics, it has nothing to do with the choice theory of either Robbins or Samuelson. It seems that reproducing some of what Ross says in the final chapter is the best way to make that case.

Ross starts out with the following general microeconomic problem and divides a proper rational choice analysis of it into three distinct steps. The first step (S1) is the following:

> Suppose you want to explain and/or predict what happens when the members of a group of one or more goal-directed systems, in causal interaction . . . pursue ends that cannot all be satisfied given available common resources that have alternative uses. In that case, use as much evidence as you . . . gather about their behavior to represent the schedule of ends pursued by each system as a . . . utility function,

[10]If an agent chooses bundle $x$ at price vector p when $x'$ was affordable, and $x \neq x'$, they have "revealed" a preference for $x$ over $x'$. Therefore, if they choose bundle $x'$ when the price vector is $p'$, it must be that their preferred bundle x was not affordable at $p'$. The WARP is thus:

$$p\,x' \leq p\,x \Rightarrow p'\,x > p'\,x'.$$

> defined as per axioms that admit of solution by simultaneously maximizing each utility function for a given allocation of resources (p. 377).

This first step involves framing the predictive/explanatory strategy in terms of goal-directed agency (note, not necessarily human agency) and identifying the constraints on the agent's action (scarcity). This is explanation/prediction of behavior in terms of maximization of stable preferences (choice functions) under scarcity—constrained optimization of a well-behaved choice function—and that is, I fully agree, what neoclassical choice theory is all about. But Ross does not stop here. He goes on to tell us that the purpose of this first step is to obtain the information necessary to *identify a particular objective function for each agent.* As he explains:

> All the first step of the RASP does is to tell us how and when to individuate economic agents. This can yield no predictions about their behavior until we have empirically justified a particular maximization function for each of them . . . For this, the economist must work in direct collaboration with the cognitive-behavioral scientist, in the way well exemplified in neuroeconomics (p. 378).

So we identify real patterns in behavior and take the intentional stance with respect to those patterns by a process of "reverse engineering" (pp. 363-69); we find explicit representations of objective functions for the relevant agents that would rationalize these observed patterns. Ross argues, it seems rightly, that neuroeconomics could play a key role in this process.

The second step (S2) essentially involves solving all of the agent's optimization problems (obtained in step one) and deriving each of their optimal choice (usually reaction) functions. In Ross' words: "Empirically identify a maximization function for each agent in the network of interactors" (pp. 363–69).

The third and final step (S3) is to specify the institutional framework of *interaction*—specifying the relevant game—and solving it:

> Identify the constraints on "G"-level games playable by the agents.
>
> Identify the specific scenario to be explained with one such game, and find that game's Nash equilibrium (pp. 363–69).

If we were to apply this RASP framework to an old-fashioned but familiar problem like Walrasian general equilibrium theory, the first step would involve rationalizing an observed pattern of prices (and quantities) by finding a set of preferences and endowments for the agents that could potentially rationalize the observed pattern. The second step would involve solving each of these individual consumer choice problems and obtaining the corresponding demand functions, and then aggregating them to get market excess demand functions. The third step would be to specify the game—in the Walrasian case the simple game of everyone taking prices as parameters until the equilibrium price vector has been found by the Walrasian *tâtonnement.* Given that the Walrasian equilibrium is also a Nash equilibrium, identification of the observed prices with the equilibrium price vector would complete the scientific exercise.

If we move beyond classical (Nash equilibrium) games to evolutionary game theory the basic framework stays the same, but things get messier—the specification of the relevant agents and the dynamics involved in reaching (some/any) equilibrium become substantially more complex. But in either case, classic/Nash or evolutionary games, these three steps together involve rationalizing the observed data by explicitly specifying the objective functions, constraints, and solution strategies that would rationalize those observations; "interests" and "strategies" are "retrospective constructs for rationalizing outcomes" (p. 365). As Ross explains the evolutionary case "the intentional-stance functionalist . . . works *backwards* from the equilibria of plausible evolutionary games to belief ascriptions that would rationalize them in light of facts about evolutionary histories and information-processing capacities" (p. 363).



With all this in mind, let us now return to Samuelson's revealed preference theory, and at this point my version of the story will emphasize different features of Samuelson's work on revealed preference than the standard story related above. The first point is that Samuelson's work on the subject falls into two distinct periods, with two different goals.[11] These two different interpretations of revealed preference theory are clear from just a quick read of his two most important papers on the subject: Samuelson (1938) and Samuelson (1948). Samuelson did not use the term "revealed preference" in the 1938 paper; the reason is of course that he was trying to purge the theory of all reference to unobservable, mentalistic, concepts such as utility and preference, and put neoclassical demand theory on a new and pristinely behaviorist foundation. As I noted elsewhere (Hands 2001, p. 68), Lavoisier did not call oxygen "revealed phlogiston"; the point was to eliminate the deficient entities, not reveal them. If Samuelson had been successful in building a new theory of consumer behavior on strictly behaviorist grounds—or rather if the profession had been willing to accept such a behaviorist theory of consumption—it would have opened the door to (at least living) non-human economic agents and perhaps mid-twentieth-century microeconomics would have looked more like psychology during this period. But then of course consumer choice theory would look totally different than it does today, and it certainly wouldn't look anything like S1–S3 of Ross' RASP since there would not be (behavioristically occult) "schedule of ends" or "utility functions" to estimate or maximize. This said, it would have been broadly positivist and it would accommodate certain non-humans, but it would not represent intentional-stance functionalism. In fact, it would not have been intentional at all; it would have been eliminativist.

Thus, Samuelson initially had the radical behaviorist's goal of purging all preference-talk from economics; so what happened? Revealed preference theory is in every contemporary microeconomics textbook and those textbooks certainly do not look behaviorist; they talk freely about both preferences and choice. To understand how revealed preference is treated in contemporary economics it is important to recognize how Samuelson originally approached the problem in the 1938 paper—not the methodological-philosophical motivations, but rather what he actually did. What he did was to *start* with demand functions themselves as the primitives of the theory and then impose WARP as a consistency condition on the price-quantity combinations associated with those demand functions. Not only was this Samuelson's approach in 1938, it is the way the theory is presented today: "A further approach to the theory of

[11]This point is made clearly in Wong (2006), a book originally published in 1978. Wong actually identified three different views, and I agree with his interpretation, but the first two will suffice here. By the way, Ross actually notes Wong's three different periods but dismisses it with the "details of these adventures don't merit attention here" (p. 100).

demand is to postulate demand functions themselves as the objects given and to impose consistency conditions on them" (McKenzie 2002, p. 22). WARP implied the same (ostensibly observable) conditions—essentially the negative semi-definiteness of Slutsky matrix—as the standard model of maximizing ordinal utility subject to a linear budget constraint. Schematically (assuming the utility function is well-behaved) the standard Slutsky results are:

$$\begin{array}{ccccc} \underset{\{x\}}{Max}\, U(x) & & x_i(p_1,p_2,\ldots,p_n,M)\ \forall \mathrm{i} & & x^T Sx \leq 0\ \forall x \neq 0 \\ & \Rightarrow & where & \Rightarrow & S_{ii} \leq 0\, \forall \mathrm{i} \\ s.t. \sum_{i=1}^{n} p_i x_i = M. & & x_i(\lambda p, \lambda M) = x_i(p,M)\ \forall \mathrm{i}\, \forall \lambda > 0 & & \mathrm{S_{ij}} = S_{ji}\, \forall \mathrm{i} \neq \mathrm{j} \end{array}$$

While revealed preference theory says:

$$\begin{array}{ccc} x_i(p_1,p_2,\ldots,p_n,M)\ \forall \mathrm{i} & & \\ where & & x^T Sx \leq 0\, \forall x \neq 0 \\ x_i(\lambda p, \lambda M) = x_i(p,M)\ \forall \mathrm{i}\, \forall \lambda > 0 & \Rightarrow & S_{ii} \leq 0\, \forall \mathrm{i} \\ \text{and} & & \\ \text{WARP} & & \end{array},$$

where replacing WARP with SARP would add the symmetry condition form the standard results to the revealed preference case. Revealed preference is just another way to get the standard Slutsky restrictions on demand functions. Notice that while revealed preference theory involves a version of "rationalization"—if certain conditions hold in the pattern of observations then the pattern could have been generated by a budget-constrained consumer maximizing a well-behaved ordinal utility function—but the theory is not consistent with S1–S3 in any other way. It starts with well-behaved demand functions and a binding budget constraint (imposed on, not found in, the "pattern") and most importantly, there is no explicit specification of the agent's choice/utility functions (no specifying "a particular maximization function for each of them" as S1 requires). Revealed preference theory says that *if* one had a "pattern" of price quantity combinations (data) that could be fit to a set of $n$ well-behaved demand functions, and those functions were homogeneous of degree zero for all prices, and also "fit" the consumer's budget constraint, and any (and every) two price-quantity vectors from the data satisfies WARP, *then* the functions so derived will also satisfy two of the other implications of standard ordinal consumer choice theory (given by the two expressions on the right hand side above). It says no more than this and it certainly has nothing whatsoever to do with "reverse engineering" or "empirically identify a maximization function for each agent in the network of interactors."

Although it seems non-controversial that this is what revealed preference theory is, this is not the idea that one would get from reading Ross' discussion of what he calls RPT. From Ross, you would get the idea that revealed preference theory is a way of getting people to *reveal their preferences*. As he puts it: "To show that RPT is useful



we must find some real structures that are usefully measured—where "usefully" means nonredundantly relevant to explanation and prediction—using coefficients and relations defined by its axioms" (p. 143) and "As early as the 1930s, economists began to take seriously Samuelson's claim to have produced a testable apparatus for determining utility functions by going into labs with live test subjects and measuring them" (p. 167).

Even if one corrects the timeline and moves to the early 1950s this was not what Samuelson's revealed preference theory was about. Determining utility functions? Labs? Coefficients? Measurement? Where and when was any of this happening during the middle of the twentieth century? There was in fact surprisingly little empirical work on revealed preference theory—Koo (1963) and Koo and Hasenkamp (1972) for example—and most of it was negative. There has been some empirical work on revealed preference theory in recent years, but even in this newer literature the results have been uneven at best (Varian 2006). But none of this is a criticism of the economists involved in revealed preference theory since they, unlike Ross, were quite clear that it was never about "determining utility functions" or "measured" coefficients. When revealed preference theory was first introduced it was radically behaviorist—concerned with eliminating utility from the analysis of economic behavior, not "determining" or "measuring" it. In its mature form WARP simply provided an alternative assumption—for many a more palatable assumption—that (when properly combined with a number of other standard assumptions) could be used to deduce the same restrictions on the Slutsky matrix as ordinal utility theory. In its mature form it is a theory that involves preference and utility, but certainly does not involve "determining," "measuring," or any of the explicit specification involved in Ross' S1. There is a contemporary game theory and experimental literature that does focus on agents "revealing" their preferences—and this is apparently Ross' point of reference—but all this literature shares with Samuelson (1948) is the term revealed preference.

Of course traditional ordinal utility theory does not require, or support, such measurement either and that is considered an epistemic virtue. The problem is that Ross seems to think that revealed preference theory changed that. It didn't. The issue for the mature version of revealed preference theory has always been to provide a different set of axioms to support exactly the same demand theory as the earlier ordinal utility theory. Given the standard assumptions, SARP and ordinal utility theory are *mathematically equivalent*; they say exactly the same thing and *neither* is about "discovering" preferences. The main point of the last one hundred years of demand theory has been that one could say all kinds of things about the relationships between prices and quantities purchased without having knowledge of the preferences or "measuring" utility. Perhaps this is a scientifically defensible position and perhaps it isn't, but in either case, it doesn't seem to be a very effective strategy to "defend" NCT by making an elaborate case in favor of a version of the theory—RASP—that is so different from the standard theory that one is ostensibly trying to defend.

None of this is to suggest that Ross might not be right about the potential for a new economic theory—one based on taking the intentional stance on observable patterns, following S1–S3, employing neuroeconomics, and such. Something like this does seem to be what going on right now in many areas of economic theory and it may turn out to be a great success. Perhaps our technology now gives us a way of measuring or determining preferences that did not exist in previous times; perhaps Ross is right about neuroeconomics providing a lot of useful new information in this regard. On a simpler note, when I go to my favorite online book seller I am always greeted with "suggestions" about books I might want to purchase. These suggestions are not random. The software has, in a sense, "revealed" my preferences; it has looked at my past purchases, modeled my preferences, and offered a solution to my utility maximization problem (it may even know my financial constraints). My claim is not that S1–S3, or Ross' general intentional-stance reverse-engineering approach, is a bad approach to economic science. I am certainly more skeptical than he is, but I am also willing to consider, and am rather intrigued by, such a program. My only point is that it is a *new approach* to economic theory—it may even be part of a new mainstream as Colander (2000) and Davis (2006) argue—but it has nothing to do with Paul Samuelson's theory of revealed preference.

This has been a long paper, but I want to make one more point about the inconsistency of Samuelson's research program and the kind of NCT represented by S1–S3. This concerns Samuelson's *Foundations* and the issues of *systematicity.* A point Ross makes repeatedly, and a point that seems to be entirely correct, is that Samuelson was systematic, and brought increased systematicity to economics (where systematic here simply means having tight priors and employing mathematics). As Ross says, and I agree, the "impulse toward the systematic" was something that "Samuelson displayed absolutely devotion to" and it was his "consistent philosophical intuition" (p. 382). I disagree though about what this means, particularly what this means within the context of Samuelson's most important work *Foundations*.

Ross (p. 103) quotes *Foundations* in detail, and correctly points out that the main goal was methodological, to develop a technique that the author believed would produce "operational" and thus empirically "meaningful" theorems, and that the entire "aim of the *Foundations* is to elucidate this method and logically unify these theorems" (p. 103). Yes, exactly, but let us actually look at the content of *Foundations* and see what this systematic strategy means in the context of that work. What Ross intends the reader to take away of course is that the systematicity in *Foundations* is the same systematicity of his version of NCT given by S1–S3, and that is simply not the case. The systematicity—and in this case we can take this to be the underlying formal mathematical structure—in Samuelson's *Foundations* differs from Ross' RASP in at least two fundamental ways. First, it does not necessarily involve optimization; optimizing models are one of the two classes of models discussed in *Foundations*, but optimization is not necessary for the systematic mathematical structure presented there (it is only part I of the book; there is also a part II). And second, the mathematical structure provides only qualitative (sign only) analysis that does not involve either explicit specification of the underlying functions *or* estimation/measurement of the relevant coefficients.

The most important thing to note is that *Foundations* was the most important single document in the neoclassical synthesis of Walrasian microeconomics and Keynesian macroeconomics. The problem is that there is no maximization (from individuals or any other "agents") going on anywhere in the Keynesian economics (or other "business cycle" models) of the 1940s. If economists were constrained to investigations of the sort allowed by Ross' S1–S3 there would have been no Keynesian revolution. In fact the most important shift in attitude between Keynes's Cambridge approach and the earlier approach of Marshall and Pigou was that it is possible to do really important work and make the world a better place without measuring individual utility (by doing macro). Samuelson's *Foundations* was in fact intended to be a systematic mathematical foundation for all of economic theory—that which involved maximization (micro) and that which involve only aggregate functional relationships that were not grounded in (or stanced by) optimizing agents (macro). Note that *Foundations* also has nothing directly to do with revealed preference theory. Revealed preference theory was about consumer behavior—behavior rather than choice originally, and a different formalism for ordinal utility theory later—but in either case it was not a "foundation" for all of economic theory.



So how did Samuelson do this? How did he develop a systematic mathematical technique that would accommodate both the microeconomic optimization-based models of the NCT sort and Keynesian macroeconomics where there was no maximization? The key was the technique he developed to do comparative statics with calculus. Throughout *Foundations* Samuelson reduces economic models to equations—equations characterizing first order conditions in optimization models and equilibrium in non-optimization-based models. For example, in a 2-variable, 1-parameter case where $x=(x_1,x_2)$ are the variables and $\beta$ is the parameter, the model might be given by the two functional equations:

$$f_1(x_1,x_2,\beta)=0,$$
$$f_2(x_1,x_2,\beta)=0.$$

If these equations "come from" an optimization problem then they would represent the first order conditions for the problem where $f$ is the objective function and $f_i=\partial f/\partial x_i$. But these equations need not refer to optimization at all. Perhaps it is a macro model with two aggregate markets—say, goods and money; then these two equations represent equilibrium conditions devoid of any optimizing implications or foundations.

Samuelson's key insight was that under the relatively weak mathematical restrictions of the implicit function theorem, such equations can be solved for the two solutions as differentiable functions of the parameters:

$$x_1^*=x_1^*(\beta) \quad \text{and} \quad x_2^*=x_2^*(\beta).$$

If these solutions are substituted back into the original equations we obtain a system that can be differentiated to obtain comparative statics results:

$$\begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 \end{bmatrix} \begin{bmatrix} \partial x_1^*/\partial\beta \\ \partial x_2^*/\partial\beta \end{bmatrix} = \begin{bmatrix} -\partial f_1/\partial\beta \\ -\partial f_2/\partial\beta \end{bmatrix}. \qquad (*)$$

These comparative statics terms are the "meaningful" theorems that are the main theme of *Foundations*. Notice two things to note about (*). First, the problem may or may not involve optimization, and the method for obtaining "meaningful" comparative statics results will differ in the two cases. If it is based on optimization then the second order conditions for the optimization problem may provide sufficient mathematical structure on the matrix on the left hand side of (*) to obtain comparative statics information about the terms $\partial x_1^*/\partial\beta$ and $\partial x_2^*/\partial\beta$ . But what if it is a problem that does not involve optimization and thus does not have the additional structure imposed by the second order conditions? How is it possible to obtain comparative statics results in this case? Enter the "correspondence principle"; if one can assume or demonstrate that the model is dynamically *stable* the stability properties will often supply the same type of restrictions as the second order conditions of optimization-based models and comparative statics results can be obtained. The first half of *Foundations* essentially presents numerous economic examples of the former (optimization case) and the second half the latter (dynamically stable case). This is the *systematic* organizing principle of *Foundations.* As Samuelson explains:

> In this study I attempt to show that there do exist meaningful theorems in diverse fields of economic affairs. . . . The first is that the conditions of equilibrium are equivalent to the maximization (minimization) of some magnitude . . .
>
> However, when we leave single economic units, the determination of unknowns is found to be unrelated to an extremum position. In even the simplest business cycle theories there is lacking symmetry in the conditions of equilibrium so that there is no possibility of directly reducing the problem to that of a maximization or minimum. Instead the dynamical properties of the system are specified, and the hypothesis is made that the system is in "stable" equilibrium or motion. By means of what I have called the *Correspondence Principle* between comparative statics and dynamics, definite *operationally meaningful* theorems can be derived from so simple a hypothesis (Samuelson 1947, p. 5).

The second thing to notice about (*) is that whether it is an optimization problem or not, the comparative statics information obtained is only *qualitative* (sign) and not *quantitative* (magnitude) in nature. One almost never has explicit functions, or specific values of those functions, to fill in the elements of the matrix on the left hand side of (*). In the case of either optimization (second order conditions) or equilibrium (stability conditions) all one generally has is information about the sign patterns of the principal minors of the relevant matrix, not the individual elements, and such restrictions will at best provide only the signs of the desired comparative statics terms.

Samuelson was devoted to systematicity, but it was not the same systematicity that Ross attributes to his version of RASP. Ross' approach requires optimization, explicit specification of the objective functions for each agent, and measurement of the specific coefficients. Samuelson's *Foundations* did not require optimization (though it was one possibility), worked with general functional forms rather than explicit representations, and produced only qualitative results (from broadly qualitative assumptions). Not only was Samuelson's approach not the same as that of Ross, if one did economics in the way that Ross recommends in S1–S3, one would never need the comparative statics technique that was Samuelson's most important contribution. Obviously one could not analyze the kind of Keynesian models that Samuelson considered so important, since specification of the underlying individual utility functions



are a necessary component of Ross' theoretical strategy. But it is not just about macro. Even in the realm of microeconomics, the novelty of Samuelson's approach is that one needs *so little information* in order to derive meaningful theorems. In particular, one does not need to know the specific objective function and particular parameter values in order to get interesting comparative statics results. Samuelson's contribution in *Foundations* (even in the micro/optimization part) is to show how economists can go a long way *without all of the information* that Ross thinks is necessary (and we are now supposed to get from neuroeconomics). Again, one can question whether Samuelson's program is the proper way to do economics, and one may or may not support Ross' general approach; the point is simply that they have nothing to do with each other (and are actually in conflict). The bottom line with Samuelson, as with Robbins, is that these programs are not Ross' program and have few if any of the characteristics that he thinks are essential for successful NCT.

## IV. CONCLUSION

The task of this paper was relatively simple: the RASP program is not, in any way, the program of either Robbins or Samuelson. In fact there are many cases where these programs clearly conflict with what Ross considers to be good scientific economics. The main features of these two research programs—both in terms of what the authors themselves said they were doing, and in terms of the features of their work that made them so influential within the economics profession—are not the same features that Ross identifies with good explanatory and predictive practice in economics (RASP).

As a final point I would like to note that no matter how exegetically well-defended my argument, Ross can easily dismiss it because he is *doing philosophy* and *not history*; as he says it "is history with a spin" (p. 75). But whether Ross himself can dismiss all this or not, I believe my points are well-taken. There are limits to rational reconstruction, and the fact Ross' Robbins-Samuelson program isn't in the work of either Robbins or Samuelson, and is, in many ways, opposed to both of their projects, shows that he has exceeded those limits. For Ross this is just "fussy intellectual history" (p. 219), but for me, and I suspect most readers of this journal, it is a serious problem. Which brings me to an obvious question. Why did Ross pursue this particular strategy? Why not just *state* what one thinks good economic science is, *state* the philosophical argumentation to defend it, *make* the case that what economists do (at least recently) is consistent with such proper practice, and then *stop*. The book would have been half as long and twice as persuasive. Why drag Robbins or Samuelson into it at all?

The best answer seems to be that the historical part of the story was necessary for the *particular way* that Ross sought to defend NCT. The philosophy of science that he endorses for (or to be more naturalist, discovers at work in) economics, is essentially a realist Dennettian version of the neo-Kantian approach he ascribes to the early positivists. There are real patterns in the phenomena and what successful science does is to maintain a unity of structure across time and variation in those patterns. The patterns are real but they change with time, technology, evolution, social interests, and a variety of other things. Scientific knowledge involves relatively invariant structures that effectively instantiate these various patterns. Thus, on this view, if economics provides objective knowledge then there must be a relatively invariant abstract structure that instantiates the constantly changing patterns of theorizing that have gone on within the history of the discipline. RASP is precisely such a structure.

If this is the role that history plays in Ross' defense of NCT then it creates a tension with avowed naturalism, and most of the critical points I have raised can be seen as instances of this tension. If naturalism involves putting science and philosophy on the same level—not privileging first philosophy as providing foundations for the assessment of science—then it seems that concepts like unifying structures and invariant formalisms should be found in, rather than imposed on, disciplinary history. One final quote from Ross is useful here to make the point. He says:

> [M]y naturalist attitude to the philosophy of science makes me sensitive to a potential charge that I am *redefining* economics from a purely abstract and conceptual point of view. This is not allowed by the rules in force here. Philosophy is led by scientific practice, not the other way around. Hence my concern to anchor my account in the actual history of economic theory, to build a reinterpretation of the foundations of that theory that does not have to demand any radical, across-the-board discontinuity or sudden paradigm shift. There is a delicate tension here (p. 215).

A delicate tension indeed: a recurrent tension between history and philosophy. It seems that one can only build an interpretation that is anchored "in the actual history of economic theory" and does not exhibit "any radical, across-the-board discontinuity or sudden paradigm shift" if in fact the "actual history" does not exhibit such changes. If the actual history does exhibit discontinuities and the story ends up being one of continuity—say one based on a RASP that remains invariant across almost one hundred years of economic theorizing—then the tension has been dissolved in favor of philosophy. Perhaps my main critical point is simply that if one is going to end up with a philosophical account then why not just start out down that path from the very beginning (and leave Robbins and Samuelson out of it)? At this point I am not exactly certain how I would assess such a (non-historical) philosophical version of Ross' project, but I do think it is extremely interesting and worthy of serious investigation. As an exercise "in the actual history of economic theory" my assessment is obviously more negative.

## REFERENCES


Backhouse, Roger and Steve Medema. 2009. "Robbins's *Essay* and the Axiomatization of Economics." *Journal of the History of Economic Thought* (Forthcoming).

Balzer, Wolfgang and Bert Hamminga, eds. 1989. *Philosophy of Economics*. Dordrecht, NL: Kluwer.

Carnap, Rudolf. 1963. "Intellectual Autobiography." In P. A. Schilpp, ed., *The Philosophy of Rudolf Carnap*. LaSalle, IL: Open Court, pp. 3–84.

Cartwright, Nancy; Jordi Cat, Lola Fleck, and Thomas Uebel. 1996. *Between Science and Politics: The Philosophy of Otto Neurath*. Cambridge: Cambridge University Press.

Churchland, Patricia S. 1986. *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, MA: MIT Press.

Churchland, Paul M. 1984. *Matter and Consciousness*. Cambridge, MA: MIT Press.

Colander, David. 2000. "The Death of Neoclassical Economics." *Journal of the History of Economic Thought* 22 (June): 127–44.





Davis, John B. 2003. *The Theory of the Individual in Economics: Identity and Value*. London: Routledge.

Davis, John B. 2006. "The Turn in Economics: Neoclassical Dominance to Mainstream Pluralism." *Journal of Institutional Economics* 2 (April): 1–20.

Debreu, Gerard. 1959. *Theory of Value*. New York: Wiley.

Dennett, Daniel. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.

Dennett, Daniel. 1991a. *Consciousness Explained*. Boston: Little Brown.

Dennett, Daniel. 1991b. "Real Patterns." *The Journal of Philosophy* 88 (January): 27–51.

Dupré, John. 1993. *The Disorder of Things*. Cambridge, MA: Harvard University Press.

Friedman, Michael. 1999. *Reconsidering Logical Positivism*. Cambridge: Cambridge University Press.

Friedman, Michael. 2001. *Dynamics of Reason*. Stanford, CA: CSLI Publications.

Friedman, Milton. 1953. "The Methodology of Positive Economics." In *Essays in Positive Economics*. Chicago: University of Chicago Press, pp. 3–43.

Glimcher, Paul W. 2003. *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*. Cambridge, MA: MIT Press.

Hands, Wade, D. 1985. "The Structuralist View of Economic Theories: A Review Essay." *Economics and Philosophy* 1 (October): 303–35.

Hands, D. Wade 2001. *Reflection Without Rules: Economic Methodology and Contemporary Science Theory*. Cambridge: Cambridge University Press.

Hands, D. Wade. 2006. "Integrability, Rationalizability, and Path-Dependency in the History of Demand Theory." In P. Mirowski and D. W. Hands, eds., *Agreement on Demand: Consumer Theory in the Twentieth Century*. Durham, NC: Duke University Press [Annual Supplement to *History of Political Economy* Volume 38], pp. 153–85.

Hausman, Daniel M. 2000. "Revealed Preference, Belief, and Game Theory." *Economics and Philosophy* 16 (April): 99–115.

Houthakker, Hendrik S. 1950. "Revealed Preference and the Utility Function." *Economica* 17 (May): 159–74.

Howson, Susan. 2004. "The Origins of Lionel Robbins's *Essay on the Nature and Significance of Economic Science*." *History of Political Economy* 36 (Fall): 413–43.

Kihlstrom, Richard, Andreu Mas-Colell, and Hugo Sonnenschein. 1976. "The Demand Theory of the Weak Axiom of Revealed Preference." *Econometrica* 44 (September): 971–78.

Koo, Anthony Y. C. 1963. "An Empirical Test of Revealed Preference Theory." *Econometrica* 31 (October): 646–64.

Koo, Anthony Y. C. and Georg Hasenkamp Georg. 1972. "Structure of Revealed Preference: Some Preliminary Evidence." *Journal of Political Economy* 80 (July–August): 724–44.

Lewin, Shira B. 1996. "Economics and Psychology: Lessons from Our Own Day From the Early Twentieth Century." *Journal of Economic Literature* 34 (September): 1293–323.

Andreu Mas-Colell Michael D. Whinston, and Jerry R. Green 1995. *Microeconomic Theory*. New York: Oxford University Press.

McKenzie, Lionel. 2002. *Classical General Equilibrium Theory*. Cambridge, MA: MIT Press.

Mirowski, Philip. 2002. *Machine Dreams: Economics Becomes a Cyborg Science*. Cambridge: Cambridge University Press.

Richardson, Alan W. 1998. *Carnap's Construction of the World: The Aufbau and the Emergence of Logical Empiricism*. Cambridge: Cambridge University Press.

Robbins, Lionel. 1935. *An Essay on the Nature & Significance of Economic Science*, second edition. London: Macmillan.

Robbins, Lionel. 1938. "Interpersonal Comparisons of Utility: A Comment." *The Economic Journal* 48 (December): 635–41.

Robbins, Lionel. 1953. "Robertson on Utility and Scope." *Economica* 20 (May): 99–111.

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

Rosenberg, Alexander. 1995. *Philosophy of Social Science*, second edition. Boulder, CO: Westview Press.

Ross, Don. 1999. *The Concept of Utility from Bentham to Game Theory.* Cape Town, South Africa: University of Cape Town Press.
Ross, Don. 2000a. "Introduction: The Dennettian Stance." In D. Ross, A. Brook, and D. Thompson, eds., *Dennett's Philosophy: A Comprehensive Assessment.* Cambridge, MA: MIT Press, pp. 1–26.
Ross, Don. 2000b. "Rainforest Realism: A Dennettian Theory of Existence." In D. Ross, A. Brook, and D. Thompson, eds., *Dennett's Philosophy: A Comprehensive Assessment.* Cambridge, MA: MIT Press, pp. 147–168.
Ross, Don. 2005. *Economic Theory and Cognitive Science: Microexplanation.* Cambridge, MA: MIT Press.
Ryle, Gilbert. 1949. *The Concept of Mind.* Chicago: University of Chicago Press.
Samuelson, Paul A. 1938. "A Note on the Pure Theory of Consumer's Behaviour." *Economica* 5 (February): 61–71.
Samuelson, Paul A. 1947. *Foundations of Economic Analysis.* Cambridge, MA: Harvard University Press.
Samuelson, Paul A. 1948 "Consumption Theory in Terms of Revealed Preference." *Economica* 15 (November): 243–53.
Sneed, Joseph D. 1971. *The Logical Structure of Mathematical Physics.* Dordrecht, NL: Reidel.
Stegmüller, Wolfgang. 1976. *The Structure and Dynamics of Theories.* New York: Springer-Verlag.
Stegmüller, Wolfgang, Wolfgang Balzer, and Wolfgang Sophn, eds. 1982. *Philosophy of Economics.* New York: Springer-Verlag.
Suppe, Frederick. 1977. *The Structure of Scientific Theories*, second edition. Urbana, IL: University of Illinois Press.
Uebel, Thomas E. 1992. *Overcoming Logical Positivism From Within: The Emergence of Neurath's Naturalism in the Vienna Circle's Protocol Sentence Debate.* Amsterdam: Editions Rodopi.
Uzawa, Hirofumi. 1971. "Preference and Rational Choice in the Theory of Consumption." In J. S. Chipman, L. Hurwicz, M. K. Richter, and H. F. Sonnenschein, eds., *Preferences, Utility, and Demand.* New York: Harcourt Brace Jovanovich, pp. 7–28.
Varian, Hal. 2006. "Revealed Preference." In M. Szenberg, L. Ramrattan, and A. A. Gottesman eds., *Samuelsonian Economics and the Twenty-First Century.* Oxford: Oxford University Press, pp. 99–115.
Wong, Stanley. 2006. *The Foundations of Paul Samuelson's Revealed Preference Theory.* London: Routledge (first edition 1978).

*Journal of the History of Economic Thought,*
*Volume 30, Number 4, December 2008*

# CHARLES BABBAGE'S INFLUENCE ON THE DEVELOPMENT OF ALFRED MARSHALL'S THEORY OF THE FIRM

BY
NEIL B. NIMAN

Recently, much attention has been directed toward the early philosophical writings of Alfred Marshall (Rafaelli 2003) in order to gain new insights into his evolutionary economics. The focus has been on the workings of the mind and how Marshall's early foray into psychology influenced his later thought. In searching for the origins of Marshall's conception of the mind, Cook (2005) makes the connection between the mind as a machine and Charles Babbage's Analytical Engine, a concept that is widely recognized as the modern precursor to what we think of today as the computer (Hyman 1982). Cook (2007) takes matters one step further and postulates that Marshall's contention that organization should be treated as a factor of production also emerged from the writings of Babbage and played an important role in much of the development of Book IV of his *Principles.*

Charles Babbage, as Lucasian Professor of Mathematics at Cambridge University, founding member of the Statistical Society, part-time geologist, cryptologist, inventor and active political reformer, was an important and influential figure in the Victorian Era. Solving the mysteries of the Analytical Machine became not only a life pursuit, but also served as a metaphor for the structure of the broader cosmos. One facet of this intellectual exercise was Babbage's attempt to develop a better understanding of "the domestic economy of the factory" contained in his 1832 work entitled, *On the Economy of Machinery and Manufactures.*

It was Babbage's extension of Adam Smith's notion of the division of labor and the subsequent development of the field of scientific management that caught Marshall's eye and subsequently had a strong influence on his own thought. In *Industry and Trade*, where Marshall looks more closely at the structure of the firm, the influence of Babbage is perhaps most readily apparent. However, by looking back from *Industry and Trade* to the *Principles of Economics*, it becomes possible to develop a better understanding of why Marshall considered organization to be a factor of production on par with labor and capital and why the success and failure of the firm rests to such a large extent on the shoulders of the business leader.

To draw these connections between Babbage and Marshall, we begin by looking at the proposition that organization is important to the success of the business

Department of Economics, University of New Hampshire, Durham, NH 03824.

ISSN 1053-8372 print; ISSN 1469-9656 online/07/04000479-490 
doi:10.1017/S1053837208000448