Ross, Don (2005) *Economic Theory and Cognitive Science*
Cambridge, MA: The MIT Press
ISBN: 9780262182461
Review by Greg Hill

Some Wittgensteinian Reservations about Neuroeconomics

What has the behavior of insects got to do with economic theory? Quite a lot according to Don Ross, who proposes a radical reconstruction of the subject in *Economic Theory and Cognitive Science*.[i] The roots of this would-be revolution in the aims, scope, and methods of economics are to be found in contemporary neuroscience, which, in Ross's view, displaces and supersedes the common sense and humdrum conception of human beings as agents who, by and large, choose effective means to their ends. Ross urges economists to jettison this "folk psychology," which turns a blind eye to the inconsistent choice-making that has been disclosed by the work of psychologists and behavioral economists. While these inconsistencies have led some economists to reject the utility-maximizing agent and the neoclassical theory used to model her decision making, Ross recommends a different course – the search for mechanisms designed by nature to maximize an end subject to constraints, that is to say, the search for entities whose behavior can be explained by the neoclassical model.

According to Ross, insects and parts of the human brain fit this description much better than whole persons do because bugs and brains have been programmed by nature to act as if they are pursuing a well-defined end in an environment where the available means are scarce. Although *a kind of* "choice making" is central to Ross's recommended scheme of inquiry, it is not the kind of choosing we have in mind when we say that

someone has decided to purchase one commodity rather than another or to bear some risks instead of others.

Ross argues that "not only should we *not* assume that economic agents are identical to actual people, but that the efficacy of economic analysis *requires* the assumption that economic agency is never even straightforwardly *coextensive with* personhood" (p. 132, emphasis in the original). To make sure the reader hasn't missed the point, Ross adds, "Let me say straight out; this is the central thesis of this book; critics take note" (p. 132).

Ross puts forward his provocative manifesto with bravado. The author's enthusiasm carries the reader through an argument that is wide-ranging and well-rehearsed. Yet, in spite of its virtues, Ross's proposed restructuring of economics suffers from a number of deep conceptual problems. What this enterprise amounts to is not a new understanding of economics, but the outline of a different subject altogether. In effect, Ross proposes a new definition of "agency," a revised idea of what it means "to choose," and therewith a new kind of "economics" – what could be called, without exaggeration, "an insect or cyborg economics" the subject matter of which bears a much more problematic relationship to the concept of *choice* than Ross allows.

I. Philosophy, Science, and Dolphin Arithmetic

"If a lion could talk we could not understand him."

Wittgenstein, *Philosophical Investigations,* part II, p. 223

Ross's deconstruction of neoclassical economics draws from a school of philosophical thinking which insists that philosophy doesn't have much to contribute to science. In this view, all questions are ultimately empirical questions, and even the truths of logic and

mathematics are subject to revision based on scientific discoveries.[ii] Practically speaking, philosophy is but the handmaiden of science, occasionally clarifying matters at the margins of investigation where scientific research hasn't yet established the way things really work. In Ross's own words, "Philosophy is to be led by scientific practice, not the other way around" (p. 215).

Although Ross affirms the unity of science, he does not contend that economics should become a branch of physics. Instead, he favors explanations which take "the intentional stance," that is, explanations which assume that the entity under investigation is a rational agent pursuing a well-defined objective. It's important to be clear that "the intentional stance" doesn't necessarily mean the vantage point of a person making a decision in pursuit of an end by considering the likely consequences of alternative courses of action. In fact, the "decision making" of a thermostat is a much better example of the sort of "intentional-stance" explanation Ross has in mind, for the thermostat is an entity the operation of which can be explained by attributing to it a capacity to "take decisions" in pursuit of a *single* goal (which cannot be said of human beings in light of the research findings of behavioral economists and cognitive scientists). Thus, we can explain and predict the "behavior" of a thermostat by supposing that it "acts" in pursuit of an "objective," converting inputs (the prevailing temperature of a room in this case) into outputs (turning a heating or cooling system on or off) depending on the difference between the actual temperature and the "desired" temperature.

Unlike scientific materialists who seek to explain human behavior in the lifeless categories of the non-biological sciences, Ross aims to extend to non-human, often inanimate, entities many concepts that belong to, or at least originate in, the "human form

of life," concepts such as "purpose," "belief," and "choice."[iii] The deepest philosophical issue surrounding Ross's enterprise is whether these concepts can be sensibly applied to such things as thermostats, robots, and insects. A skeptic of this endeavor will point out that a thermostat is a *mechanism* the functioning of which has no meaning for the thermostat itself. A thermostat functions according to the laws of nature, which are external to it. By contrast, the concepts that orient human activity, such as "promising," "voting," and "making a bank deposit," are *internal* to these activities; they *orient and guide* behavior, whereas the parts of a thermostat and, more generally, the bodies studied by physicists and neuroscientists, neither "follow," nor "obey," the laws of nature.[iv] Concepts such as "bank deposits," "lines of credit," and "bankruptcy" are not theoretical abstractions invented by social scientists to explain regular patterns of behavior. Rather, these concepts belong to the institution of "banking" in the absence of which there would be no such things as "deposits," "lines of credit," and "bankruptcies."

Some contemporary thinkers who emphasize the distinctive character of human behavior also draw a distinction between the aims and methods of philosophy, on the one hand, and those of science, on the other.[v] According to this understanding, philosophy deals with *concepts* and their elucidation, for example, the concepts of "reason" and "choice" and what's involved in "making a choice for a reason," whereas science is concerned with empirical questions such as what causes bodies to move in the ways they do. Ross (and the philosophers he draws upon) reject the distinction between philosophy and science, as well as the corresponding distinction between conceptual and empirical truths. In elaborating his view, Ross asserts the following, "Surely it cannot be a *conceptual* truth that dolphins, or any possible aliens – or cyborgs – could not grasp the

proposition that '2 + 2 = 4' unless they have the same types of neural states as people" (p. 40, emphasis in original). This preemptive strike against critics who might raise conceptual objections to Ross's enterprise misfires because such critics (including this one) are not likely to insist that having "the same types of neural states as people" is the relevant criterion for deciding whether dolphins, aliens, or cyborgs could "grasp the proposition that '2 + 2 = 4.'" Whether a dolphin could grasp that 2 + 2 = 4 (let alone "*the proposition* that '2 + 2 = 4'") is a question worth pursuing because it sheds light on the claim that is central to Ross's whole enterprise, namely Ross's contention that the issue of whether insects, parts of the brain, and other non-persons can act as agents in pursuit of ends is not a conceptual, but an empirical, question (p. 222).

Suppose we undertake a scientific study to determine whether dolphins can be trained to add 2 + 2. We might start by placing two beach balls in front of the dolphin, and then two more, rewarding the dolphin whenever it immediately barks four times. Suppose that after a few weeks or months of training the dolphin will bark four times whenever two objects *of any kind* are placed alongside two other objects *of any kind.* A conceptual question now arises: does this performance by the dolphin *count as* "grasping that 2 + 2 = 4"? The answer is by no means straightforward. Typically, human beings grasp this notion after we have learned to count by using our fingers. We don't learn this particular sum in isolation, but along with other sums. We give verbal answers to questions of arithmetic or write them down, and our mistakes are *corrected*, not merely discouraged by "negative reinforcement." When we are learning basic arithmetic, we are taught the difference between addition and subtraction and the way in which these operations are applied for practical purposes. Bearing this background in mind, let us

now ask whether a dolphin "grasps that 2 + 2 = 4" if it reliably barks four times whenever two objects are set alongside two others, *even if this is the only sum it has "mastered"?*

Ross's suggestion to the contrary notwithstanding, the ground-floor conceptual issue isn't whether a dolphin (without a human neural network) could grasp that 2 + 2 = 4, but *what counts as grasping that 2 + 2 = 4*, and this *is* a conceptual, not an empirical, question. Human beings learn simple sums in the context of many other things: a mother tells her child he can only have two pieces of candy; the child complains that his older sister has three pieces; a neighbor holds up two fingers and asks the child if he has more than "this many" pieces of candy; and so on. It's against this background that we ascertain whether someone, usually a child (which is also important in arriving at this judgment), has "grasped that 2 + 2 = 4." In the case of the dolphin, this background is lacking and, as a consequence, we lack the criteria that are available to us in judging whether a person "grasps that 2 + 2 = 4." Perhaps Ross would conclude that a dolphin which barked four times whenever two objects were set alongside two others had indeed "grasped that 2 + 2 = 4." Would he also insist that the dolphin's *grasping* of this sum is *the same* as the child's grasping of it?

Let's press a little harder because Ross's example of "dolphin addition" is much more plausible than many other kinds of behavior that might be attributed to dolphins (not to mention to insects, thermostats, and neural networks). Is it a conceptual, or a scientific, question whether a dolphin could come to regret something it did many years ago (or that a dolphin could make choices in accordance with a "minimax regret" decision rule)? We know what it is for a human being to regret something he did many years ago. He may express it in recounting the story of a business venture that turned out

badly. We may hear it in his voice and see it in his facial expressions as he laments his failure to help a friend who needed his assistance a long time ago. What do we look for in a dolphin's behavior to see whether the dolphin regrets something it did (or didn't do!) a long time ago? I don't rule out the possibility that dolphin life could evolve in such a manner that dolphins could come to "have regrets." But if it did, dolphin life would have to become much more like human life (or vice versa), otherwise *we couldn't grasp the regrets expressed by dolphins*. If the life of dolphins (or of humans) did evolve in ways that brought new commonalities to our two forms of life, then scientists might investigate the conditions that made "dolphin regret" possible, but – and this is the essential point – an investigation of this kind presupposes a set of criteria for what *counts as having regrets*, and this is not a task for scientific investigation, but rather is presupposed by it.

Ross's response to this line of criticism is worth quoting at length because the success of his venture depends on it:

> "Primarily conceptual arguments are unreliable guides to empirical facts. Such arguments must rely for their force on intuitions; but intuitions are (biologically and/or culturally) evolved devices for helping people form expectations in particular environments and with respect to salient and important objects and events in those environments, classified for practical purposes. It is thus no surprise that intuitions systematically mislead us if we rely on them when seeking general truths holding over nonparochially selected samples of reality" (pp. 123-124).

The force of this argument obviously depends on what Ross means by "intuitions." Does the argument about the possibilities of dolphin experience outlined

above appeal to “intuitions”? I certainly didn’t claim that my intuition, or our intuition, or most people’s intuition is that dolphins couldn’t possibly grasp that 2 + 2 = 4. Rather, I pointed out that the criteria for ascertaining whether a dolphin grasps this sum are not readily at hand. We know what it means for a human being to grasp that 2 + 2 = 4. We do not know, and Ross offers us no help in this regard, what criteria we should apply to determine whether a dolphin has grasped that 2 + 2 = 4. This claim about the absence of relevant criteria doesn’t mean we lack an “intuition” about the possibility of dolphin addition, but that we lack the standards on the basis of which we could ascertain whether a dolphin had grasped that 2 + 2 = 4, not to mention “*the proposition* that 2 + 2 = 4.”

Ross’s argument against this kind of conceptual critique involves three claims: 1) conceptual arguments rely on intuitions; 2) intuitions are “biologically and/or culturally evolved devices for helping people form expectations in particular environments”; and 3) such intuitions will mislead us in seeking “general truths holding over nonparochially selected samples of reality.” If, by “intuition,” Ross means something like our “gut feeling” or first impression, as in, “I just can’t imagine a dolphin adding 2 + 2,” then there’s merit to his argument. But if he means by “intuition” the fact that we agree *as a matter of course* that “2 + 2 = 4,” or that “a pentagon has five sides,” or that “the moon is not made of cheese,” in other words, if Ross aims to rule out any appeal to the fact that we agree in a great many simple judgments, then he is going to find it very difficult to explain how we can have any concepts at all, whether these concepts belong to our everyday world or to the research activities of neuroscientists. Science, itself, is only possible because scientists routinely have the same “intuition” about such mundane matters as the readings displayed on their measuring instruments.

II. The Subject Matter of Economics

> "Only of a living human being and what resembles (behaves like) a living human being can one say, it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious."
>
> Wittgenstein, *Philosophical Investigations*, sec. 281

Although Ross embraces Lionel Robbins's conception of economics as the study of choice among means with alternative uses, he sees no reason to limit the class of choice-making agents to human beings, nor, in the end, any reason even to admit "whole persons" into the category of choice-making agents (p. 87). For Robbins, the paradigmatic economic agent is the person who chooses X rather than Y because X is a less costly means of achieving the person's objective, Z. In this simple conception, there's an agent with a goal and a set of beliefs about the most efficient means of attaining it. But whereas most economists (having given the matter little thought) implicitly assume that goals, and beliefs about alternative means, are to be found, metaphorically at least, "in the mind" of the agent, Ross rejects this understanding as a misguided piece of "folk psychology," which implies that people act on the basis of "psychologically real, causally active, representations of objects" (p. 135).

Against this mistaken notion from "folk psychology," Ross urges an "intentional-stance functionalist interpretation" of human behavior in which goals and beliefs are not to be found in the minds or brains of decision-making agents, but "only ascribed to subjects for the purpose of describing regularities in their behavior" (p. 135). Thus, we are advised to assume that the entities under consideration act as if they are pursuing an objective and have rational beliefs about the most efficient means of achieving it.

Finally, the criterion for choosing the best belief to attribute to a person (or a thermostat) is whether the belief "predicts behavioral patterns with maximal efficiency" (p. 61).

Once beliefs have thus been stripped of their "mindedness," human beings lose their unique status as agents pursuing ends in light of the alternatives they imagine, and the proper subject of economics comes into view. Economics is not about people who choose among scarce resources in pursuit of their objectives. Rather, the proper subjects of economics are "systems in nature whose behavior can be nonredundantly predicted and explained through comparing *available* trade-offs in search of the *best* (most efficient) trade-offs" (p. 117, emphasis in the original). How do we ascertain the goal of such a system? Since Ross includes insects and parts of the human brain among these systems, Robbins's method of introspection will not do the job. Instead, Ross urges us to proceed along the lines recommended by Samuelson – assume behavior reveals the "preference ordering" of the subject under investigation – but he then adds a surprising twist: do not regard these "subjects" as coextensive with human beings (p. 116). If there are "systems in nature" the behavior of which can be explained and predicted in terms of optimization, then, Ross concludes, "some Samuelsonian systems describe real patterns, and the science of these systems constitutes 'economics' as approximately defined by Robbins" (p. 117).

Ross's insistence to the contrary notwithstanding, this notion of "explaining" and "predicting" behavior by "comparing trade-offs" actually bears very little resemblance to Robbins's understanding of economics. The research program Ross proposes is based on a misconception of both human choice-making and nature's evolutionary processes. To begin with, intentional stance functionalism gives a misleading account of the way in

which we understand human behavior. Consider first, as an illustrative contrast, the "behavior" of a thermostat. If we hear the mechanism inside a thermostat "click," we may assume "the intentional stance" and ascribe to the thermostat the "belief" that the temperature has crossed a certain threshold, and infer that the thermostat has "decided" to turn the heater on or off depending on the temperature "it desires to maintain." Do we proceed in an analogous fashion when considering, for example, a person who is signing a business contract? No. When we characterize certain "bodily movements" as signing a contract, we have already invested these "movements" with a kind of mindedness, that is, with intention. We *assume* the person signing the contract believes she is entering into a contractual agreement. We do not "ascribe" this belief in order to explain her movement of the pen across a piece of paper. Such a belief is already implicit in her actions, otherwise we wouldn't say she was "signing a contract."

When we explain an action in terms of a person's ends or desires and her beliefs about the best, or most appropriate, means available to pursue them, we are not advancing a *theory* about her action. It's not as if, in trying to understand someone's actions, we are confronted with "raw data" for which we must develop an explanatory theory. The interpretation of someone's behavior is not typically a two-step process that begins with a perception of some bodily movements and acoustic phenomena and then proceeds to a consideration of whatever beliefs might best explain these sounds and movements, allowing us to predict their future course. Many actions are not even identifiable as particular kinds of action unless they are already regarded as issuing from a particular set of desires and beliefs. This is true of making promises, deceiving others, committing many kinds of crimes, buying insurance, and applying for a job. We make

judgments about such intentions often, and in large part, based on the circumstances in which these actions take place, the behavior leading up to the actions in question, the social norms relevant to these performances, the character of the people involved if we know them, and considerations of this nature.

A person living in a society that lacks the practices constitutive of contractual relationships cannot enter into a business contract; nor can a baby, a thermostat, or a part of one's brain. The vassals who inhabited Europe's medieval communities could not enter into employment contracts because they lacked the freedom to do so; babies cannot undertake contractual commitments because they lack the requisite understanding and legal status; and neither thermostats nor parts of the brain can enter into contracts for the same reason that pianos can't get married – it's a nonsensical notion.

This is only in part to reiterate the point that concepts come with criteria on the basis of which we ascertain whether some piece of behavior counts as signing a contract, fulfilling one's contractual obligations, breaking a contract, and so forth. I also want to make the related point that we cannot meaningfully attribute beliefs to people who, given their social and historical circumstances, could not possibly hold these beliefs *whether or not such attributions had some predictive value*. It makes no sense to attribute a belief in "the effectiveness of monetary policy" to someone who has no understanding of central banking even if the attribution of such a belief were somehow useful in predicting this person's behavior. The criterion for correct belief attribution is not predictive power, but whether attributing the belief in question to this particular person makes sense in light of her own circumstances and experience.

It may be objected that while "signing a contract" is deeply embedded in a social and legal setting, "comparing available trade-offs in search of the most efficient trade-offs" is not similarly embedded and, therefore, economists don't face the same obstacles in extending this notion to non-persons. But what is gained in terms of understanding, explanation, or predictive power, by regarding a thermostat as an "agent" that "believes" the temperature has risen to "its desired level" and therefore "decides" to turn the heater off? If someone knows how a thermostat works, is her understanding deepened, or her predictions made more accurate, if she construes the mechanical movements involved as "decisions"? If the thermostat fails to maintain the temperature we desire, will it be helpful to criticize its failure, to give it further instructions, to encourage it to try harder? These are, after all, the kinds of things we do when an agent has failed to perform a particular task.

Ross's redefinition of economics also stumbles because he has assimilated an evolutionary conception of "trade-offs" to the notion of "trade-offs" that economists (and ordinary people) invoke to explain the alternatives facing someone *making a choice*. "Evolutionary trade-offs" are relevant when, for example, we imagine that an animal with "extra" body fat would have enhanced survival prospects in winter when food is scarce, but reduced survival prospects in summer when speed and agility are important. Here we can speak of a "trade-off" associated with extra body fat that is advantageous in the winter, but disadvantageous in the summer. And we might employ the notion of such a trade-off in explaining why a particular species evolved in one way rather than another. But, that said, it's nonsensical to add that Mother Nature actually made a "choice" in this

matter, “weighing” the costs and benefits of extra body fat before arriving at her “decision.”

This notion of trade-offs, that is, alternatives with advantages and disadvantages, is at home in evolutionary game theory where strategies flourish depending on the game’s payoff structure and the strategies employed by, or embodied in, other players. One can meaningfully speak of a trade-off in evolutionary game theory where a particular strategy, X, does well against an alternative strategy, Y, but not against strategy Z. In the process of “evolution,” that is to say, after many rounds of play, one or more of these strategies may predominate. But this does not mean, nor does it make sense to say, that Mother Nature, or the process of natural selection, or repeated plays of the game have “evaluated” the trade-offs involved and “chosen” the winning strategies. Neither Mother Nature, nor natural selection, nor repeated plays of a game is an agent who assesses alternatives, makes decisions, and judges the results, at least not if “assess,” “decide,” and “judge” are to retain their present meaning.

Evolutionary game theory, insofar as it involves fixed strategies, differs from those game-theoretic conceptions in which the players, themselves, choose the strategies they play. For example, a player in an Assurance Game may choose between “cooperating” or “defecting,” the best strategy being to “cooperate” if one believes the other player will “cooperate” and to “defect” if one believes otherwise. This is a genuine choice in which each player makes an assessment of the other player’s intention and, on the basis of that assessment, decides whether to cooperate or defect. By contrast, Mother Nature, in an evolutionary game, makes no “assessment” of the “players’ intentions,” nor any “choice” regarding the strategy that will prevail over a long sequence of interactions.

Where the players actually choose a strategy, we may ask about their expectations regarding the likely choices of other players; or about their estimate of the payoffs attached to different outcomes; or whether, once the outcome of the game is known, they regret the strategy they had chosen. To ask Mother Nature about her reasoning concerning such matters is pointless: there is no answer in prospect because Mother Nature neither chooses nor explains the reasons "she had in mind" in favoring one strategy over another, for Mother Nature is not an agent to "whom" it makes sense to attribute choices and reasons.

In light of the difficulties we encountered in trying to articulate what would count as "a dolphin grasping the proposition that 2 + 2 = 4," not to mention the conditions under which it would make sense to say that a dolphin had "expressed regret for something done a long time ago," the shortcomings of Ross's conception of "agency," "choice," and related terms should, by now, be clear. In simplest terms, the criteria on the basis of which we call something a "choice" or a "decision," and the criteria we draw upon in judging whether such-and-such is an "agent" or, to suggest but one of many contrasts, a "principal," cannot be readily extended to thermostats, insects, and parts of the brain, *at least not if the words "agent" and "choice" are to retain their customary meaning*. Can we sensibly call something an agent if it can't explain or justify <u>any</u> of its "choices" or "actions"? Is anything illuminated if we throw out the distinction between a change in temperature that causes a thermostat to turn on a heater and a change of view that leads an investor to increase her holdings of Euro-denominated bonds?

Ross concedes that the concept of "agency" is closely connected with the concepts of "action," "responsibility," and "moral judgment," that we hold people

responsible for the actions they take rather than for events over which they have no control. However, Ross thinks he can sever the conceptual links between "agency," "personhood," and "moral responsibility," because, in his view, these notions are only tied together in our misguided "folk psychology." According to Ross, the "folk" equate "agency" with behavior that originates "within" persons, specifically within their "minds." Yet, he insists, this fundamental proposition of "folk psychology" is simply false; behavioral scientists haven't found any evidence that there are "causally efficacious states" of the brain (p. 238). Thus, Ross concludes, the "language games" of folk psychology revolving around "responsibility," "praise," and "blame," are only sustainable because words like "agency" are used very loosely, and the "folk" who use them are unaware of the voluminous evidence that falsifies the "folk" understanding of human behavior and its causes. For Ross, "what enables the folk who play language games involving agency to go on is merely a network of false beliefs about their own psychology" (p. 238).

Here Ross mistakes the *grammar* of agency and responsibility, that is, their conceptual interconnections and the rules governing the use of these concepts, for a (mistaken) *theory* of human action which holds that agency is possible because "people have causally efficacious states in their brains corresponding to isolable beliefs and desires" (p. 238). In fact, we do not analyze a person's brain in order to decide whether she was responsible for something (except in rare cases in which, for example, a person has suffered brain damage that might bear on the attribution of responsibility). Whether someone should be held accountable for a particular act depends upon a variety of considerations: the social norms relevant to the action under review, what the person

knew or should have known about the circumstances surrounding the action and its possible consequences, whether the person could have done otherwise, whether she was under duress, and so on. None of these conditions requires as a necessary premise that "people have causally efficacious states in their brains corresponding to isolable beliefs and desires," and therefore Ross is simply mistaken to hold that the "language games of agency and responsibility" only persist because of misguided views about the relationship between brains and behavior.

III. From Economics to Cognitive Science

> "I want to say: 'If someone could see the mental process of expectation, he would necessarily be seeing <u>what</u> was expected' . . . he would not have to <u>infer</u> it from the process he perceived."
>
> Wittgenstein, *Philosophical Investigations*, sec. 452-53, original emphasis

Ross wishes to export the study of decision making from economics to cognitive science. Instead of asking people "what they think they're doing and why," which, according to Ross, is "an inherently unreliable method" for understanding the causes of behavior, "one needs to know, instead, what actual objective function governs people's behavior as a result of their natural and causal histories" (p. 172). This is, to say the least, an ambitious research program. We are familiar with explanations that "go deeper" than the reasons people give for their actions. We invoke causes of various kinds to explain the behavior of alcoholics, drug addicts, people with obsessive-compulsive disorders, even "panic selling" on Wall Street. But we typically turn to causal explanations to account for abnormal, or irrational, behavior. If a person says he's going to take the bus to the theater to avoid the high cost of parking downtown; or that he's going to reduce his bond

holdings because he thinks interest rates will soon rise; or that he's going to increase his rate of saving because his daughter has her heart set on Harvard, we don't, except in unusual circumstances, find it necessary to search for the "actual objective function" that "governs" this person's behavior.

The trouble with Ross's proposed investigation of the "actual objective function [that] governs people's behavior as a result of their natural and causal histories" is that it does not take seriously the notion of "acting for a reason." If I were to condense all my criticisms of this book into a single sentence, it would be that Ross believes the causal model of explanation is the only model worthy of consideration, ignoring altogether the *logical* relationship between actions and reasons. It is because Ross thinks only in terms of "cause and effect" that he first characterizes reasons as supposedly "causally effective representations," and then dismisses this account of acting for reasons, as though reasons, if they are to have any bearing on human behavior, must be located inside the agent's head, giving her intentions a kind of "causal push" to get them going. The explanation of an action in terms of reasons does not carry us behind the scene of the action to its causal source. When we ask why a person chose a particular course of action, we are not looking for causes, but for motives and, sometimes, for justifications. Ross, having ruled out the very idea of acting for a reason, lacks the conceptual resources necessary to distinguish between a reflex action in which a person's hand jerks in response to a stimulus such an electric shock and an intentional action in which a person raises her hand in order to ask a question.

Consider the implications of Ross's view for the discipline he holds in highest regard – science itself. Are scientific theories *judged* on the basis of their explanatory

and predictive power, or have theories such as quantum mechanics merely gained ascendancy in the scientific community because they maximize some "objective function" that "governs" the behavior of scientists? If a scientist says she accepts Darwin's theory of evolution because of the mountain of evidence in favor of it, should we regard her explanation as "inherently unreliable" and wait for a better one from cognitive scientists studying the "real" bases of human behavior? And on what grounds, if any, are we to judge the theories advanced by cognitive scientists?

Ross argues that we're caused to hold certain beliefs because they enhance our prospects for survival. Thus, when crossing the street, our chances of surviving are much better if we believe we'll be injured if hit by a car; that a car is approaching at high speed; and that we'd better move quickly if we don't want to get run over. But Ross is blind to the fact that we hold many beliefs because *we have reasons for holding them*. One could make a case for the "holding-beliefs-for-reasons-view" by pointing out that it gives a plausible explanation of many beliefs Ross wishes to explain in terms of causes. Why do I believe a car will hit me if I don't move out of the way? Because I can see it coming towards me! Why do I think I'll be seriously injured or die if the car hits me? Because I've seen the results of car accidents. Why do I believe I must move quickly? Because the car is rapidly approaching, and I don't want to be hit by it. By regarding the relationship between the world and our thoughts as one of cause and effect, Ross misses *the object of belief*, and, as a consequence, the way in which the content of our experience can provide reasons for the (not always correct) beliefs we come to hold. A car coming towards me can be both the object of my belief, as well as the reason for holding it, regardless of how I respond to the approaching car (consider the person

planning suicide). By conceiving of our experience as the effect of external stimuli, Ross loses hold of the idea that experience can provide a justification, a reason, for our beliefs.

IV. Conclusion: Neoclassical Economics – A Theory in Search of a Subject

Neoclassical economics has evolved an elegant mathematical model to explain the behavior of rational agents and the consequences of their interaction. Meanwhile, psychologists and behavioral economists have found many examples in which the behavior of real-life human beings has turned out to be inconsistent with the neoclassical model. Although some economists have reacted to this state of affairs by relinquishing the rational utility-maximizing decision-maker in favor of a more complex choice-making agent, Ross's approach is to search for entities that do behave as if they were rational decision makers pursuing a well-defined objective in the face of constraints. In other words, rather than modifying the rational choice model in light of its explanatory shortcomings, Ross sets out to find organisms and mechanisms whose behavior can be explained by the rational choice model, which turn out to be such things as insects, thermostats, and parts of the brain.

There is, however, an alternative explanation of both the shortcomings of the neoclassical model in economics and the success of similar (optimization) models in explaining the behavior of non-persons. Although Ross credits Keynes for liberating economists from the need to build every model upon "microfoundations," it was Keynes's recognition of uncertainty's central place in the human predicament that revealed the limits of economic models which assume that meaningful probabilities can be attached to outcomes, so that rational agents can select the most efficient means to their ends. The movement of (most) *things* can be easily predicted because the laws of

nature push them toward a unique equilibrium; a marble placed on a curved surface will come to rest at its lowest point. But human beings are not marbles; we have a conception of our circumstances and can act with an objective in mind. And, yet, while the power to imagine alternative courses of action releases us from the web of predictable causation, the resulting stream of spontaneous choices denies us the possibility of forming the kind of rational expectations the neoclassical model implicitly assumes. It is not that human beings are bad statisticians, which may well be true, but that the object of our predictions – the choices of others and the consequences of these choices – will not submit to the discipline we seek to impose upon it.

greg.hill@seattle.gov

[i] Don Ross, *Economic Theory and Cognitive Science* (Cambridge, MA: The MIT Press, 2005).

[ii] See W. V. O. Quine, "Two Dogmas of Empiricism," in Quine, *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1953), pp. 20-46.

[iii] I am using Wittgenstein's phrase, "form of life." See Ludwig Wittgenstein, *Philosophical Investigations*, translated by G. E. M. Anscombe (Oxford: Blackwell, 1953), part II, 174. See also Greg Hill, "Solidarity, Objectivity, and the Human Form of Life," *Critical Review* (1998) Vol. 11, No. 4: 555-80.

[iv] See Peter Winch, *The Idea of a Social Science and its Relation to Philosophy* (Routledge & Kegan Paul: New York, 1958).

[v] See P. M. S. Hacker, *Wittgenstein's Place in Twentieth-Century Analytic Philosophy* (Oxford: Blackwell, 1996).