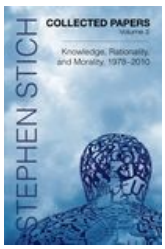

Collected Papers, Volume 2: Knowledge, Rationality, and Morality, 1978-2010

Stephen Stich

https://doi.org/10.1093/acprof:oso/9780199733477.001.0001

Published: 2012 **Online ISBN:** 9780199949823 **Print ISBN:** 9780199733477

CHAPTER

6 Naturalizing Epistemology: Quine, Simon, and the Prospects for Pragmatism

Stephen Stich

https://doi.org/10.1093/acprof:oso/9780199733477.003.0006 Pages 99–112

Published: August 2012

Abstract

In recent years there has been a great deal of discussion about the prospects of developing a "naturalized epistemology," though different authors tend to interpret this label in quite different ways. This chapter sketches three projects that might lay claim to the "naturalized epistemology" label, and argues that they are not all equally attractive. Indeed, the first of the three—the one attributed to Quine—is simply incoherent. There is no way we could get what we want from an epistemological theory by pursuing the project Quine proposes. The second project is a naturalized version of reliabilism. While this project is not fatally flawed in the way that Quine's is, the sort of theory this project would yield is much less interesting than might at first be thought. The third project is located squarely in the pragmatist tradition. One of the claims made for this project is that if it can be pursued successfully the results will be both more interesting and more useful than the results that might emerge from the reliabilist project. A second is that there is some reason to suppose that it can be pursued successfully. It is argued that for over a decade one version of the project has been pursued with considerable success by Herbert Simon and his co-workers in their ongoing attempt to simulate scientific reasoning. The final section offers a few thoughts on the various paths Simon's project, and pragmatist naturalized epistemology, might follow in the future.

Keywords: naturalized epistemology, Quine, reliabilism, Herbert Simon, scientific reasoning

Subject: Philosophy, Epistemology, Moral Philosophy, Philosophy of Mind

1. Introduction

In recent years there has been a great deal of discussion about the prospects of developing a "naturalized epistemology," though different authors tend to interpret this label in quite different ways.[1] One goal of this paper is to sketch three projects that might lay claim to the "naturalized epistemology" label, and to argue that they are not all equally attractive. Indeed, I'll maintain that the first of the three—the one I'll attribute to Quine—is simply incoherent. There is no way we could get what we want from an epistemological theory by pursuing the project Quine proposes. The second project on my list is a naturalized version of reliabilism. This project is not fatally flawed in the way that Quine's is. However, it's my contention that the sort of theory this project would yield is much less interesting than might at first be thought.

The third project I'll consider is located squarely in the pragmatist tradition. One of the claims I'll make for this project is that if it can be pursued successfully the results will be both more interesting and more useful than the results that might emerge from the reliabilist project. A second claim I'll make for it is that there is some reason to suppose that it *can* be pursued successfully. Indeed, I will argue that for over a decade one p. 100 version of the project *has* been pursued with considerable success by Herbert Simon and his ↳ co-workers in their ongoing attempt to simulate scientific reasoning. In the final section of the paper, I will offer a few thoughts on the various paths Simon's project, and pragmatist naturalized epistemology, might follow in the future.

Before I get on to any of this, however, I had best begin by locating the sort of naturalistic epistemology that I'll be considering in philosophical space. To do this I'll need to say something about how I conceive of epistemology, and to distinguish two rather different ideas on what "naturalizing" might come to. Much of traditional epistemology, and much of contemporary epistemology as well, can be viewed as pursuing one of three distinct though interrelated projects. One of these projects is the assessment of strategies of belief formation and belief revision. Those pursuing this project try to say which ways of building and rebuilding our doxastic house are good ways, which are poor ways, and why. A fair amount of Descartes' epistemological writing falls under this heading, as does much of Bacon's best-known work. It is also a central concern in the work of more recent writers like Mill, Carnap, and Goodman. A second traditional project aims to provide a definition or characterization of knowledge, explaining how knowledge differs from mere true opinion, as well as from ignorance and error. A third project has as its goal the refutation of the skeptic—the real or imagined opponent who claims that we can't have knowledge or certainty or some other putatively valuable epistemological commodity.[2] Although these three projects are obviously intertwined in various ways, my focus in this paper will be exclusively on the first of the three. The branch of epistemology whose "naturalization" I'm concerned with here is the branch that attempts to evaluate strategies of belief formation.

Let me turn now to "naturalizing." What would it be to "naturalize" epistemology? There are, I think, two rather different answers that might be given here. I'll call one of them Strong Naturalism and the other Weak Naturalism. What the answers share is the central idea that empirical science has an important role to play in epistemology—that epistemological questions can be investigated and resolved using the methods of the natural or social sciences. The issue over which Strong Naturalism and Weak Naturalism divide is the *extent* to which science can resolve epistemological questions. Strong Naturalism maintains that *all* legitimate epistemological questions are scientific questions, and thus that epistemology can be reduced to or replaced by science. Weak Naturalism, by contrast, claims only that *some* epistemological questions can be resolved by science. According to Weak Naturalism there are some legitimate epistemological questions that are *not* scientific questions and cannot be resolved by scientific research. The sort of epistemological pragmatism that I'll be advocating in this paper is a version of Weak Naturalism. It claims that while some epistemological questions can be resolved by doing science, there is at least one quite fundamental epistemological issue that science cannot settle.

p. 101

2. Quine(?)'s Version of Strong Naturalism

The most widely discussed proposal for naturalizing epistemology is the one sketched by Quine in "Epistemology Naturalized" (1969b) and a number of other essays (1969c, 1975). According to Quine,

> [Naturalized epistemology] studies a natural phenomenon, viz., a physical human subject. This human subject is accorded a certain experimentally controlled input—certain patterns of irradiation in assorted frequencies, for instance—and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. The relation between the meager input and the torrential output is a relation that we are prompted to study for somewhat the same reasons that always prompted epistemology; namely, in order to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence. (1969b, 82–83).
>
> The stimulation of his sensory receptors is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world. Why not just see how this construction really proceeds? Why not settle for psychology? (1969b, 75–76)

There are various ways in which this Quinean proposal might be interpreted. On one reading, Quine is proposing that psychological questions can replace traditional epistemological questions—that instead of asking: How *ought* we to go about forming beliefs and building theories on the basis of evidence? we should ask: How do people actually go about it? And that the answer to this latter, purely psychological question will tell us what we've really wanted to know all along in epistemology. It will tell us "how evidence relates to theory." I'm not at all sure that this is the best interpretation of Quine.[3] What I am sure of is that many people do interpret Quine in this way. I am also sure that on this interpretation, Quine's project is a non-starter.

To see why, let us begin by asking *which* "physical human subject" or subjects Quine is proposing that we study. Quine doesn't say. Perhaps this is because he supposes that it doesn't much matter, since we're all very much alike. But that is simply not the case.

Consider, for example, those "physical human subjects" who suffer from Capgras syndrome. These people typically believe that some person close to them has been kidnapped and replaced by a duplicate who looks and behaves almost exactly the same as the original. Some people afflicted with Capgras come to believe that the replacement is not human at all; rather it is a robot with electrical and mechanical components
p. 102 inside. There have even been a few cases reported in which the Capgras sufferer attempted to prove ↳ that the "duplicate" was a robot by attacking it with an axe or a knife in order to expose the wires and transistors concealed beneath the "skin." Unfortunately, not even the sight of the quite real wounds and severed limbs that result from these attacks suffice to persuade Capgras patients that the "duplicate" is real.[4] Now for a Capgras patient, as much as for the rest of us, "the stimulation of his sensory receptors is all the evidence [he] has had to go on, ultimately, in arriving at his picture of the world." And psychology might well explore "how this construction really proceeds." But surely this process is *not* one that "we are prompted to study for the same reasons that always prompted epistemology." For what epistemologists want to know is not how "evidence relates to theory" in any arbitrary human subject. Rather they want to know how evidence relates to theory in subjects who do a good job of relating them. Among the many actual and possible ways in which evidence might relate to theories, which are the *good* ways and which are the bad ones? That is the question that "has always prompted epistemology." And the sort of study that Quine seems to be proposing cannot possibly answer it.

People suffering from Capgras syndrome are, of course, pathological cases. But much the same point can be made about perfectly normal subjects. During the last two decades cognitive psychologists have lavished

considerable attention on the study of how normal subjects go about the business of inference and belief revision. Some of the best-known findings in this area indicate that in lots of cases people relate evidence to theory in ways that seem normatively dubious to put it mildly (Nisbett and Ross 1980; Kahneman, Slovic, and Tversky 1982). More recent work has shown that there are significant interpersonal differences in reasoning strategies, some of which can be related to prior education and training (Fong, Krantz, and Nisbett, 1986; Nisbett, Fong, Lehman, and Cheng 1987). The Quinean naturalized epistemologist can explore in detail the various ways in which different people construct their "picture of the world" on the basis of the evidence available to them. But he has no way of ranking these quite different strategies for building world descriptions; he has no way of determining which are better and which are worse. And since the Quinean naturalized epistemologist can provide no normative advice whatever, it is more than a little implausible to claim that his questions and projects can replace those of traditional epistemology. We can't "settle for psychology" because psychology tells us how people *do* reason; it does not (indeed cannot) tell us how they *should*.[5]

3. Reliabilism: Evaluating Reasoning by Studying Reasoners Who Are Good at Forming True Beliefs

The problem with Quine's proposal is that it doesn't tell us *whose* psychology to "settle for." But once this
p. 103
has been noted, there is an obvious proposal for avoiding the problem. ↳ If someone wants to improve her chess game, she would be well advised to use the chess strategies that good chess players use. Similarly, if someone wants to improve her reasoning, she would be well advised to use the reasoning strategies that good reasoners use. So rather than studying just anyone, the naturalized epistemologist can focus on those people who do a good job of reasoning. If we can characterize the reasoning strategies that good reasoners employ, then we will have a descriptive theory that has some normative clout.[6]

This, of course, leads directly to another problem. How do we select the people whose reasoning strategies we are going to study? How do we tell the good reasoners from the bad ones? Here there is at least one answer that clearly will *not* do. We can't select people to study by first determining the reasoning strategies that various people use, and then confining our attention to those who use good ones. For that would require that we already know which strategies are good ones; we would be trying to pull ourselves up by our own bootstraps. However, as the analogy with chess suggests, there is a very different way to proceed. We identify good chess players by looking at the consequences of their strategies—the good players are the ones who win, and the good strategies are the ones that good players use. So we might try to identify good reasoners by looking at the outcome of the reasoning. But this proposal raises further questions: Which "outcomes" should we look at, and how should we assess them? What counts as "winning" in epistemology?

One seemingly natural way to proceed here is to focus on *truth*. Reasoning, as Quine stresses, produces "descriptions of the ... world and its history." A bit less behavioristically, we might say that reasoning produces *theories* that the reasoner comes to *believe*. Some of those theories are true, others are not. And, as the example of the Capgras sufferer's belief makes abundantly clear, false theories can lead to disastrous consequences. So perhaps what we should do is locate reasoners who do a good job at forming *true* beliefs, and try to discover what strategies of reasoning they employ. This project has an obvious affinity with the reliabilist tradition in epistemology. According to reliabilists, *truth* is a quite basic cognitive virtue, and beliefs are justified if they are produced by a belief-forming strategy that generally yields true beliefs. So it would be entirely in order for a naturalistically inclined reliabilist to propose that reasoning strategies should be evaluated by their success in producing true beliefs.[7]

It might be thought that this proposal suffers from something like the same sort of circularity that scuttled the proposal scouted two paragraphs back since we can't identify ↳ reasoners who do a good job at producing true theories unless we already know how to distinguish true theories from false ones. On my view, this charge of circularity can't be sustained. There is no overt circularity in the strategy that's been sketched, and the only "covert" circularity lurking is completely benign, and is to be found in all other accounts of how to tell good reasoning strategies from bad ones. However, I won't pause to set out the arguments rebutting the charge of circularity, since Goldman has already done a fine job of it.[8]

But while this reliabilist project is not viciously circular, it is, I think, much less appealing than might at first be thought. In support of this claim, I'll offer two considerations, one of which I've defended at length elsewhere. The project at hand proposes to distinguish good reasoning strategies from bad ones on the basis of how well they do at producing true beliefs. But, one might well ask, what's so good about having true beliefs? Why should having true beliefs be taken to be a fundamental goal of cognition? One's answer here must, of course, depend on what one takes true beliefs to be. If, along with Richard Rorty, one thinks that true beliefs are just those that one's community will not challenge when one expresses them, then it is not at all clear why one should want to have true beliefs, unless one values saying what one thinks while avoiding confrontation.[9]

I am not an advocate of Rorty's account of truth, however. On the account I favor, beliefs are mental states of a certain sort that are mapped to propositions (or content sentences) by an intuitively sanctioned "interpretation function." Roughly speaking, the proposition to which a belief-like mental state is mapped may be thought of as its truth condition. The true beliefs are those that are mapped by this function to true propositions; the false beliefs are those that are mapped to false propositions. However, it is my contention that the intuitively sanctioned function that determines truth conditions—the one that maps beliefs to propositions—is both arbitrary and idiosyncratic. There are lots of other functions mapping the same class of mental states to propositions in quite different ways. And these alternative functions assign different (albeit counter-intuitive) truth conditions. The class of beliefs mapped to true propositions by these counter-intuitive functions may be slightly different, or very different from the class of beliefs mapped to true propositions by the intuitive function. So, using the counter-intuitive functions we can define classes of beliefs that might be labeled TRUE* beliefs, TRUE** beliefs, and so on. A TRUE* belief is just one that is mapped to a true proposition by a counter-intuitive mapping function. Yet many of the alternative functions are no more arbitrary or idiosyncratic than the intuitively sanctioned function. Indeed, the only special feature that the intuitively sanctioned function has is that it is the one we happened to have been bequeathed by our language and culture. If all of this is right, then it is hard ↳ to see why we should prefer a system of reasoning that typically yields true beliefs over a system that typically yields TRUE* beliefs. The details on all of this, and the supporting arguments, have been set out elsewhere (Stich 1990, ch. 5; 1991a; 1991b). Since there is not space enough to reconstruct them here, let me offer a rather different sort of argument to challenge the idea that good reasoning strategies are those that typically yield true beliefs.

If one wants to play excellent chess, one would be well advised to use the strategies used by the best players in their best games. Of course, it *may* be possible to do even better than the best players of the past. One can always hope. But surely a good first step would be to figure out the strategies that the best players were using at the height of their power. For, barring cosmic accident, those are likely to be very good strategies indeed. Now suppose we were to try to apply this approach not to chess strategies but to reasoning strategies. Whose reasoning would we study?

Here opinions might differ, of course. But I suspect that most of us would have the great figures of the history of science high on our list. Aristotle, Newton, Dalton, Mendel—these are some of the names that would be on my list of Grand Masters at the "game" of reasoning. If one is a reliabilist, however, there is something quite odd about this list. For in each case the theories for which the thinker is best known, the theories they produced at the height of their cognitive powers, have turned out not to be true. Nor is this an

idiosyncratic feature of this particular collection of thinkers. It is a commonplace observation in the history of science that much of the best work of many of the best scientific thinkers of the past has turned out to be mistaken. In some cases historical figures seem to be getting "closer" to the truth than their predecessors. But in other cases they seem to be getting further away. And in many cases this notoriously obscure notion of "closer to the truth" seems to make little sense.

The conclusion that I would draw here is that if we adopt the strategy of locating good reasoners by assessing the *truth* of their best products, we will end up studying the wrong class of thinkers. For some of the best examples of human reasoning that we know of do not typically end up producing true theories. If we want to know how to do a good job of reasoning—if we want to be able to do it the way Newton did it—then we had better not focus our attention exclusively on thinkers who got the right answer.

4. Pragmatism: There Are No Special Cognitive Goals or Virtues

The project sketched in the previous section might be thought of as having two parts. The first part was entirely normative. It was claimed that truth was a quite special cognitive virtue, and that achieving true beliefs was the goal in terms of which strategies of reasoning should be evaluated. The second part was empirical. Having decided that good cognition was cognition that produced true belief, we try to identify people who excel by that measure, and then study the way they go about the business of reasoning. Using the terminology suggested in Section 1, the project is a version of Weak Naturalism. Science, broadly
p. 106 construed, can tell us which reasoners do a good job at producing true beliefs, ↳ and what strategies of reasoning they exploit. But science can't either confirm or disconfirm the initial normative step. Science can't tell us by what standard strategies of reasoning *should* be evaluated. The critique of the project that I offered in the previous section was aimed entirely at the normative component. It is, I argued, far from obvious that producing true beliefs is the standard against which strategies of reasoning should be measured.

But if truth is not to be the standard in epistemology, what is? The answer that I favor is one that plays a central role in the pragmatist tradition. For pragmatists, there are *no* special cognitive or epistemological values. There are just *values*. Reasoning, inquiry, and cognition are viewed as tools that we use in an effort to achieve what we value. And like any other tools, they are to be assessed by determining how good a job they do at achieving what we value. So on the pragmatist view, the good cognitive strategies for a person to use are those that are likely to lead to the states of affairs that he or she finds intrinsically valuable. This is, of course, a thoroughly relativistic account of good reasoning. For if two people have significantly different intrinsic values, then it may well turn out that a strategy of reasoning that is good for one may be quite poor for the other. There is, in the pragmatist tradition, a certain tendency to down play or even deny the epistemic relativism to which pragmatism leads. But on my view this failure of nerve is a great mistake. Relativism in the evaluation of reasoning strategies is no more worrisome than relativism in the evaluation of diets or investment strategies or exercise programs. The fact that different strategies of reasoning may be good for different people is a fact of life that pragmatists should accept with equanimity.[10]

As I envision it, the pragmatist project for assessing reasoning strategies proceeds as follows. First, we must determine which goal or goals are of interest for the assessment at hand. We must decide what it is that we want our reasoning to achieve. This step, of course, is fundamentally normative. Empirical inquiry may be of help in making the decision, but science alone will not tell you what your goals are. Thus the pragmatist's project, like the reliabilist's, is a version of Weak Naturalism. The second step is to locate people who have done a good job at achieving the goal or goals selected. The third step—and typically it is here that most of the hard work comes in—is to discover the strategies of reasoning and inquiry that these successful subjects have used in achieving the specified goal. Just as in the case of chess, the expectation is that if we can

discover the strategies used by those who have done a good job at achieving the goals we value, these will be good strategies for us to use as well. But we need not assume that they are the best possible strategies. It may well be that once we gain some understanding of the strategies used by people who have excelled in achieving the specified goals, we may find ways of improving on their strategies. Exploring the possibility of improving on the actual strategies of successful cognitive agents is the fourth step in the pragmatist project.

p. 107

5. Herbert Simon's Computational Pragmatism

The pragmatist project sketched in the previous section is of a piece with the epistemological theory I defended in *The Fragmentation of Reason*. Shortly after that book was completed I was delighted to discover that for more than two decades Herbert Simon and his colleagues had been hard at work on a project that had all the essential features of the one I have proposed. They had long been practicing what I had only recently started to preach.[11]

Simon's project is an ambitious research program in artificial intelligence. He characterizes the project, rather provocatively, as an attempt to construct a "logic of scientific discovery." The "logic" that Simon seeks would be an explicit set of principles for reasoning and the conduct of inquiry which, when followed systematically, will result in the production of good scientific hypotheses and theories. As is generally the case in artificial intelligence, the principles must be explicit enough to be programmed on a computer. Simon and his co-workers don't propose to construct their logic of discovery by relying on *a priori* principles or philosophical arguments about how science should proceed; their approach is much more empirical. To figure out how to produce good scientific theories, they study and try to simulate what good scientists do. In some ways their project is quite similar to "expert systems" studies in AI. The initial goal is to produce a computational simulation of the reasoning of people who are "experts" at doing science.

Though Simon does not stress the point, he acknowledges that a largely parallel project might be undertaken with the goal of simulating the reasoning of some other class of "experts." We might, for example, focus on the reasoning of people who have done outstanding work in history, or in literary criticism, or in theology. In some of these cases (or all of them) we might end up with pretty much the same principles of reasoning. But then again, we might not. It might well turn out that different strategies of reasoning work best in different domains. The choice of which group of reasoners to study—and ultimately, the choice of which strategy to use in one's own reasoning—is the initial normative step in Simon's pragmatic project.

Having decided that the reasoning he wants to study is the sort that leads to success in science, the second step in Simon's project is to identify people who have achieved scientific success. As a practical matter, of course, this is easy enough. There is a fair amount of agreement on who the great scientists of the past have been. But when pressed to provide some justification for the scientists he selects, Simon (only half jokingly) p. 108 suggests the ↳ following way to "operationalize" the choice: Go to the library and get a collection of the most widely used basic textbooks in various fields. Then sit down and make a list of the people whose pictures appear in the textbooks. Those are the people whose reasoning we should study. Though I rather doubt that Simon has ever actually done this, the joke makes a serious point. The criterion of success that Simon is using is not the *truth* of the theories that various scientists produce. To be a successful scientist, as Simon construes the notion, is to be famous enough to get one's picture in the textbooks.

With a list of successful scientists at hand, the really challenging part of Simon's project can begin. The goal is to build a computational simulation of the cognitive processes that led successful scientists to their most celebrated discoveries. To do this, a fair amount of historical information is required, since optimally the input to the simulation should include as much as can be discovered about the data available to the scientist who is the target of the simulation, along with information about the received theories and background

assumptions that the scientist was likely to bring to the project. As with other efforts at cognitive simulation, there is a variety of evidence that can be used to confirm or disconfirm the simulation as it develops. First, of course, the simulation must end up producing the same law or theory that the target scientist produced. Second, the simulation should go through intermediate steps parallel to those that the scientist went through in the course of making his or her discovery. In some cases, laboratory notebooks and other historical evidence provide a quite rich portrait of the inferential steps (and missteps) that the target scientist made along the way. But in most cases the details of the scientist's reasoning are at best very sketchy. In an effort to generate more data against which the simulation can be tested, Simon and his co-workers have used laboratory studies of problem solving and "rediscovery" in which talented students are asked to come up with a law or theory that will capture a set of data, where the data provided are at least roughly similar to the data available to the target scientist. While they are working, the students are asked to "think out loud" and explain the various steps they make. The problems are often very hard ones, and relatively few students succeed. But the protocols generated by the successful students can be used as another source of data against which simulation programs can be tested (Kulkarni and Simon 1988; Dunbar 1989; Qin and Simon 1990).

It should be stressed that there is no *a priori* guarantee that Simon's research program will be successful. There is a long tradition which insists that scientific creativity, indeed all creativity, is a deeply mysterious process, far beyond the reach of computational theories. And even if we don't accept the mystery theory of creativity, it is entirely possible that efforts to simulate the reasoning which led one or another important scientist to a great discovery will fail. The only way to silence these concerns is to deliver the goods. It is also possible that while each individual scientist's reasoning can be simulated successfully, each case is different. There might be no interesting regularities that all cases of successful scientific reasoning share. Perhaps successful scientific reasoning is discipline specific, and different strategies of reasoning are
p. 109 successful in different disciplines. Worse still, it might ↳ turn out that no two successful scientists exploit the same strategies. Styles of successful reasoning might be entirely idiosyncratic. Having noted these concerns, however, I should also note that it doesn't look like things *are* turning out this way. While there is still lots of work to be done, Simon and his group have produced impressive simulations of Kepler's discovery of his third law, Krebs' discovery of the urea cycle, and a variety of other important scientific discoveries. While some of the heuristics used in these simulations are specific to a particular scientific domain, none are specific to a particular problem, and many appear to be domain independent (Kulkarni and Simon 1990, Section 5). So, though the jury is still out, I think it is entirely reasonable to view Simon's successes to date as an excellent beginning on the sort of pragmatist naturalization of epistemology that I advocated in the previous section. In the final section of this paper, I want to consider some of the ways in which Simon-style pragmatist projects may develop in the future.

6. Beyond History's Best: Future Projects for Naturalistic Pragmatism

The project of simulating successful scientific reasoning is the one that has preoccupied Simon and his co-workers up until now. However, once some substantial success has been achieved along these lines—and it is my reading of the situation that we are now at just about that stage—it becomes possible to explore some new and very exciting territory. As the historical record indicates, important discoveries are often slow in coming and they frequently involve steps that later come to be seen as unnecessary or unfruitful. To the extent that simulations like Simon's have as their goal understanding the details of the psychological process that lead to discoveries, it is, of course, a virtue if they explore blind alleys just where the scientists they were modeling did. However, if we want a normative rather than a descriptive theory of discovery, it is no particular virtue to mimic the mistaken steps and wasted efforts of gifted scientists. Thus rather than aiming to describe the cognitive strategies of gifted scientists, we might aspire to *improve* on those strategies. By tinkering with the program—or, more interestingly, by developing a substantive theory of how and why they work—we may well be able to design programs that do *better* than real people, including very gifted and highly trained people, be they important historical figures or clever students in laboratory studies of reasoning. I think that to a certain extent this sort of tinkering and theory-driven improvement is already a part of Simon's project, though it is often not clearly separated from the process of modeling actual discovery. The process of improvement can be pursued along several rather different lines. What distinguishes them is the sort of constraints that the computational model takes to be important. In the remaining pages of this paper I want to sketch some of the *constraints* that might be imposed or ignored, and consider the sorts of projects that might ensue.

A first division turns on how much importance we attach to the idea that normative rules and strategies of p. 110 reasoning have to be usable by human beings. To the extent that we ↳ take that constraint seriously, we will not propose strategies of reasoning that are difficult or impossible for a human cognitive system. Our normative theory will respect the limitations imposed by human psychology and human hardware. A natural label for this project might be *Human Epistemology*. Of course, the more our normative theory of Human Epistemology respects the limits and idiosyncrasies of human cognition, the closer it will resemble the descriptive theory of good reasoning. But there is no reason to think that the two will collapse. For it may well be the case that there are readily learnable and readily usable strategies of reasoning that would improve on those that were in fact used in the "exemplary" cases of scientific discovery. In order to pursue Human Epistemology in a serious way we will need detailed information about the nature and the rigidity of constraints on human cognition. And the only way to get this information is to do the relevant empirical work. This is yet another way in which the sort of naturalized epistemology that I am advocating requires input from empirical science.[12]

What happens if we are not much concerned with constraining our epistemological system by taking account of the facts of human cognition? We aren't free of all constraints, since the commitment to construct theories of scientific discovery that are explicit enough to be *programmable* imposes its own constraints. The theories we build must be implementable with available hardware and available software. But, of course, there are lots of things that available systems can do quite easily that human brains cannot do at all. So if we are prepared to ignore the facts about human cognition, we are likely to get a very different family of normative theories of scientific discovery. In recent work, Clark Glymour has introduced the term *Android Epistemology*, and I think that would be an ideal label to borrow for normative theories like these.

If there were more space available, I would spend it exploring the prospects for Android Epistemology. For it seems to me that they are very exciting prospects indeed. What is slowly emerging from the work of Simon's group, and from the work of other groups focusing on related problems, is, in effect, a *technology of discovery*. We are beginning to see the development of artifacts that can discover useful laws, useful concepts, and useful theories. It is, of course, impossible to know how successful these efforts will

ultimately be. But I, for one, would not be at all surprised if future historians viewed this work as a major juncture in human intellectual history.

Let me return to the domain of Human Epistemology. For there is one more distinction that needs to be drawn here. Once again the notion of constraints provides a convenient way to draw the distinction. One of the facts about real human cognizers is that they are embedded in a social context. They get information and support from other people, they compete with others in various ways, and their work is judged by others. p. 111 Many of the rewards for their efforts come from the surrounding society as the result of ↳ these judgments. In building our *Normative Human Epistemology* we may choose to take account of these factors or to ignore them.

In their work to date, Simon and his colleagues have largely chosen to ignore these social constraints. And for good reason. Things are complicated enough already, without trying to see how well our simulations do when competing with other simulations in a complex social environment. Nonetheless, I think there may ultimately be a great deal to learn by taking the social constraints seriously, and exploring what we might label *Social Epistemology*. For example, Philip Kitcher (1990) has recently tried to show that the likely payoff of pursuing long shots in science—the expected utility of working hard to defend implausible theories—depends in important ways on the distribution of intellectual labor in the rest of the community. I think there is reason to hope that if we take seriously the idea of building epistemological theories for socially embedded cognitive agents we may begin to find ways in which the organization of the inquiring community itself may be improved. We may find better ways to fund research, channel intellectual effort, deal with dishonesty, and distribute rewards. As a pragmatist, I can think of no finer future for epistemology.[13]

References

Buchanan, B. 1983. Mechanizing the Search for Explanatory Hypotheses. In P. Asquith and T. Nichols, eds., *PSA 1982*, vol. 2. East Lansing, MI: Philosophy of Science Association.
Google Scholar Google Preview WorldCat COPAC

Dunbar, K. 1989. Scientific Reasoning Strategies in a Simulated Molecular Genetics Environment. In *Proceedings of the Eleventh Annual Meeting of the Cognitive Science Society*. Ann Arbor, MI: Erlbaum.

Foerstl, H. 1990. Capgras' delusion. *Comprehensive Psychiatry* 31: 447–49.
WorldCat

Fong, G., D. Krantz, and R. Nisbett. 1986. The Effects of Statistical Training on Thinking About Everyday Problems. *Cognitive Psychology* 18: 253–92. 10.1016/0010-0285(86)90001-0
WorldCat Crossref

Goldman, A. 1986. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
Google Scholar Google Preview WorldCat COPAC

Kahneman, D., P. Slovic, and A. Tversky, eds. 1982. *Judgment Under Uncertainty*. Cambridge: Cambridge University Press.
Google Scholar Google Preview WorldCat COPAC

Kim, J. 1988. What Is "Naturalized Epistemology"? *Philosophical Perspectives* 2: 381–405. 10.2307/2214082
WorldCat Crossref

Kitcher, P. 1990. The Division of Cognitive Labor. *Journal of Philosophy* 87: 5–22. 10.2307/2026796
WorldCat Crossref

Kitcher, P. 1992. The Naturalists Return. *Philosophical Review* 101: 53–114. 10.2307/2185044
WorldCat Crossref

Kornblith, H., ed. 1985a. *Naturalizing Epistemology*. Cambridge, MA: MIT Press.
Google Scholar Google Preview WorldCat COPAC

Kornblith, H. 1985b. *What Is Naturalistic Epistemology?* In Kornblith 1985a, 1–13.
Google Scholar Google Preview WorldCat COPAC

p. 112 Kulkarni, D., and H. Simon. 1988. The Processes of Scientific Discovery: The Strategy of Experimentation. *Cognitive Science* 12: 139–75. 10.1207/s15516709cog1202_1
WorldCat Crossref

Kulkarni, D., and H. Simon. 1990. *Experimentation in Machine Discovery.* In Shrager and Langley 1990a.
Google Scholar Google Preview WorldCat COPAC

Langley, P., H. Simon, G. Bradshaw, and J. Zytkow. 1987. *Scientific Discovery: Computational Explorations of the Creative Processes*. Cambridge, MA: MIT Press.
Google Scholar Google Preview WorldCat COPAC

Nisbett, R., ed. 1993. *Rules for Reasoning*. Hillsdale, NJ: Erlbaum.
Google Scholar Google Preview WorldCat COPAC

Nisbett, R., G. Fong, D. Lehman, and P. Cheng. 1987. Teaching Reasoning. *Science* 238: 625–31. 10.1126/science.3672116
WorldCat Crossref

Nisbett, R., and L. Ross. 1980. *Human Inference*. Englewood Cliffs, NJ: Prentice-Hall.
Google Scholar Google Preview WorldCat COPAC

Qin, Y., and H. Simon. 1990. Laboratory Replication of Scientific Discovery Processes. *Cognitive Science* 14: 281–312. 10.1207/s15516709cog1402_4
WorldCat Crossref

Quine, W. 1969a. *Ontological Relativity and Other Essays*. New York: Columbia University Press.
Google Scholar Google Preview WorldCat COPAC

Quine, W. 1969b. *Epistemology Naturalized.* In Quine 1969a, 69–90. Reprinted in Kornblith 1985a.
Google Scholar Google Preview WorldCat COPAC

Quine, W. 1969c. *Natural Kinds.* In Quine 1969a, 114–38. Reprinted in Kornblith 1985a.
Google Scholar Google Preview WorldCat COPAC

Quine, W. 1975. The Nature of Natural Knowledge. In S. Guttenplan, ed., *Mind and Language*. Oxford: Clarendon Press.
Google Scholar Google Preview WorldCat COPAC

Rorty, R. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
Google Scholar Google Preview WorldCat COPAC

Rorty, R. 1982. *Consequences of Pragmatism*. Minneapolis: University of Minnesota Press.
Google Scholar Google Preview WorldCat COPAC

Rorty, R. 1988. Representation, Social Practice, and Truth. *Philosophical Studies* 54: 215–28. 10.1007/BF00354514
WorldCat Crossref

Shrager, J., and P. Langley, eds. 1990a. *Computational Models of Scientific Discovery and Theory Formation*. San Mateo, CA: Morgan Kaufmann Publishers.
Google Scholar Google Preview WorldCat COPAC

Shrager, J., and P. Langley. 1990b. *Computational Approaches to Scientific Discovery.* In Shrager and Langley 1990a.
Google Scholar Google Preview WorldCat COPAC

Simon, H. 1966. Scientific Discovery and the Psychology of Problem Solving. In R. Colodny, ed., *Mind and Cosmos: Essays in Contemporary Science and Philosophy*. Pittsburgh: University of Pittsburgh Press.
Google Scholar Google Preview WorldCat COPAC

Simon, H. 1973. Does Scientific Discovery Have a Logic? *Philosophy of Science* 40: 471–80. 10.1086/288559
WorldCat Crossref

Stich, S. 1990. *The Fragmentation of Reason*. Cambridge, MA: MIT Press.
Google Scholar Google Preview WorldCat COPAC

Stich, S. 1991a. *The Fragmentation of Reason*—Precis of Two Chapters. *Philosophy and Phenomenological Research* 51: 179–83. 10.2307/2107833
Crossref

Stich, S. 1991b. Evaluating Cognitive Strategies: A Reply to Cohen, Goldman, Harman and Lycan. *Philosophy and Phenomenological Research* 51: 207–213. 10.2307/2107838
WorldCat Crossref

Thagard, P. 1988. *Computational Philosophy of Science*. Cambridge, MA: MIT Press.
Google Scholar Google Preview WorldCat COPAC

Zytkow, J., and H. Simon. 1988. Normative Systems of Discovery and Logic of Search. *Synthese* 74: 65–90. 10.1007/BF00869619
WorldCat Crossref

Notes

1. For useful discussions of these various interpretations see Kornblith 1985b and Kitcher 1992.
2. For more on these three projects, see Stich 1990, 1–4.
3. Indeed, in earlier drafts of this paper I attributed the view to someone called "Quine(?)" as a way of emphasizing my uncertainty about the interpretation. But that device survives only in the title of this section; it gets old very quickly.
4. Foerstl 1990. I am grateful to Lynn Stephens for guiding me to the literature on Capgras syndrome.
5. For a similar critique of Quine, see Kim 1988.
6. A number of people have suggested to me that this strategy of studying the reasoning of people who are good at it is what Quine actually had in mind. I find relatively little in Quine's writing to support this interpretation. But I do not pretend to be a serious scholar on such matters. If those who know Quine's work better than I decide that this is what he really intended, I'll be delighted. I can use all the support I can get.
7. The sort of naturalized reliabilism that I am sketching bears an obvious similarity to the psychologically sophisticated reliabilism championed by Alvin Goldman. See, for example, Goldman 1986.
8. See Goldman 1986, 116–21. In Stich 1990, Section 6.3, I have added a few of my own bells and whistles to Goldman's arguments.
9. This is, of course, no more than a caricature of Rorty's view. The full view defies easy summary. See Rorty 1979, Ch. 8; 1982, xiii–xlvii; 1988.
10. For more on pragmatism and relativism, see Stich 1990, Section 6.2.
11. The literature in this area is extensive and growing quickly. While Simon is clearly a seminal figure, many others have done important work. In much of what follows, "Simon" should be read as shorthand for "Simon and his co-workers." Perhaps the best place to get an overview of Simon's work in this area is in Langley, Simon, Bradshaw, and Zytkow 1987. For a review of more recent work see Shrager and Langley 1990b and the other essays in Shrager and Langley 1990a. Other useful sources include Simon 1966, 1973; Buchanan 1983; Kulkarni and Simon 1988, 1990; Zytkow and Simon 1988; Thagard 1988.
12. For some interesting studies aimed at discovering how much plasticity there is in human reasoning, see the papers in Nisbett 1993, Part VI.
13. Earlier versions of this paper were presented at the Conference on Methods at the New School for Social Research, the Southern Society for Philosophy and Psychology, the Australasian Association for Philosophy and the conference on Philosophy and Cognitive Science at the University of Birmingham. I am grateful to the audiences at all these meetings for many helpful suggestions. Thanks are also due to Peter Klein for extended comments on the penultimate version of the paper, and to Paul Lodge for help in preparing the final version of the manuscript.