from *Philosophy and Psychology*, eds. R. Kitchener and W. O'Donohue, London: Sage Publications, 1996, pp. 19-32.

Psychology, Naturalized Epistemology and Rationality

Harold I. Brown

1. Descriptions and Norms

The concept of rationality has both a descriptive and a normative component. The side descriptive is exemplified by Aristotle's account of human beings as rational animals--creatures that have a rational part of the soul in addition to the vegetative and sensitive parts that we share with other animals. The normative side of rationality stands out when we recognize that although human beings have a capacity to be rational, there are situations in which we fail to behave rationally. Our ability to recognize failures of rationality and to criticize individuals for such failures requires norms for the proper exercise of our rational capacities. Now once we note this dual nature of the concept of reason, it becomes especially important to understand the relation between the two components. Yet, to raise this question is to raise one of philosophy's most difficult traditional problems: the relation between norms and descriptions. Historically this problem has most commonly arisen in the context of ethics but it appears whenever we reflect on norms, including epistemic norms. Indeed, an account of the relation between norms and description is particularly pressing for proponents of a naturalistic epistemology. I want to develop these issues in somewhat greater detail.

In ethics one aspect of the relation between norms and descriptions was first made explicit in Hume's dictum that we cannot deduce ought from is:

> In every system of morality, which I have hitherto met with, I have always remark'd, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the usual copulations of propositions, is, and is not, I meet with no proposition that is not connected with an ought, or an ought not. This change is imperceptible; but is, however, of the last consequence. For as this ought, or ought not, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. (Hume 1978, p. 469)

Note that Hume is making a narrow logical point in this passage, but the point is extremely important because Hume does not allow for any legitimate inferences other than deductions--Hume is famous for arguing that there are no legitimate inductive inferences.[1] Note also that the key consequence he draws from his logical point is that our normative ethical decisions have no cognitive foundation. One possible response to Hume is that there is more to cognition than is captured in deduction, and this is a point that I shall pursue in the course of this paper. But for the moment it is important to be clear that a wider issue is at stake than just a point about deduction. For our central aim in seeking norms is to be able to decide the right thing to do in various situations, and we need norms exactly because people do not always do the right thing. But how are we to determine the norms for correct behavior? Hume's point is that we cannot arrive at an account of correct behavior solely from a description of the way people do in fact behave, even when this description is supplemented by an account of the nature of the universe and by deductive logic.

The argument thus far suggests that we must be able to establish norms quite independently of an account of how people actually behave; but this is only half the story. A short journey down the historical path--from Hume to Kant--will bring us to the other half. Also in the context of ethical theory Kant maintained that ought implies can. The implication referred to is deductive implication and Kant's aim was to provide a principle to undercut

claims that we are unable to carry out the dictates of morality. But "ought implies can" is logically equivalent to "not-can implies not-ought," that is, that we are not obliged to do what is beyond our power. I shall be concerned here with Kant's dictum in this latter form and, in particular, with its significance for the case of epistemic norms. The effect of Kant's thesis is to counter-balance Hume's in a rather interesting way. For while Hume's thesis entails that we cannot derive epistemic norms solely from a description of cognitive abilities, Kant's thesis entails that we cannot establish epistemic norms without an account of cognitive abilities. If we attempt to proceed a priori we may well end up with norms that have no legitimate force for human beings because they make demands on us that we cannot possibly fulfill. In other words, we need an account of the appropriate epistemic norms for human beings; an account of what is normatively rational requires a prior account of what is rational in the descriptive sense--an account of what cognitive abilities human beings have available.

Note especially that what is normatively rational may vary for different beings depending on their cognitive abilities. So, what is normatively rational for adult human beings may be quite different from what is rational for young children, or properly programmed computers, or perhaps for Martians. Moreover, we might even have to recognize different rational demands on different people in different circumstances. I will not be able to explore all of these variations here, but the general point is crucial; some terminology due to Reiner (1993) will help clarify what is at stake. Reiner notes that the term "rational" has two quite different contraries--arational and irrational--and this verbal distinction marks the exact distinction that we need. An arational being lacks the ability to behave rationally; we can think of a descriptive scale capturing different degrees of rationality with arationality as its zero point. A scale running from irrational to rational will be a normative scale, one that seeks to capture what an individual ought to do given the available capacities. But we cannot begin to locate an individual on the normative scale until we have located that individual on the descriptive scale; we cannot decide which norms are appropriate for us until we have determined where we are on the descriptive scale.

The conclusion we have just arrived at directly contradicts a dominant view in mainstream contemporary epistemology. On this view, normative epistemology must proceed independently of any of the descriptive subjects exactly because people can put forth descriptive claims that are incorrect. Thus, it is argued, we must already have a set of epistemic norms in order to decide which proposed descriptions are correct. This contradiction is important because there are powerful arguments on both sides. I will attempt to resolve the contradiction later in this paper, but it will be useful to set the stage by pursuing some further historical reflections. In doing so we will find that descriptive epistemology has played a much greater role in the development of normative epistemology than many contemporaries epistemologists have acknowledged. Moreover, we shall see that descriptive theses have been fundamental not only in providing the basis for specific norms, but also in determining what kinds of norms are appropriate for human knowers.

2. Some Problems of Epistemology

We can begin by noting an obvious feature of human cognition: we are not omniscient but we are capable of learning a great deal on a wide number of topics. Although the point is obvious, it has played a central role in determining the problems of epistemology and their possible solutions. If we were omniscient then tasks such as finding the appropriate rules for evaluating knowledge claims would have no point--nor would there by a point to this quest if we were incapable of expanding the body of our knowledge. Moreover, specific features of human beings have also determined how we pursue the epistemologist's tasks. To take another obvious example, much effort has been expended in trying to understand how reliable our senses are, how we can improve their reliability, and how we can use information derived from sense perception to evaluate hypotheses that go beyond the limits of what we have

already sensed. If we were organisms of a very different kind, this particular problem-cluster would not have played a central role in epistemology. Indeed, the central role of the "problem of perception" in philosophy is one of the characteristic features of the modern period that begins (roughly) with Descartes. But this new problematic emerged largely because of changes in our understanding of the actual nature of human sense perception. It will be worth our while to sketch some of these changes.

According to most widely held view of sense perception in the later middle ages, our senses typically show us the physical world exactly as it is. The basic picture of how this comes about is founded in Aristotelian metaphysics. Briefly, Aristotle conceived of a physical object as consisting of a form imposed on a material substrate; all of the properties that characterize an object are included in the form. One crucial motivation for this view is that it allow us to understand the sense in which, for example, two distinct dollar bills are items of the same kind: we have a single form that is instantiated in two different bits of matter. Now just as the same form can occur in different material objects, that form can also be instantiated in a human mind, and this is what happens when we perceive a physical object. Aristotle provided an account of the human mind that made sense of this proposal. One crucial feature of this account is that the mind has no internal structure of its own that enters into the perceptual process and "distorts" the form of the perceived object. Several medieval thinkers sought to provide a more detailed account of how the form gets imprinted on the mind but I will not pursue these details here (see Lindberg 1976 for discussion). I want, rather, to consider some developments that led philosophers to question the Aristotelian account of perception. I note in passing that a challenge to the view of the mind as lacking any internal structure also develops and becomes central with Kant, but I will limit discussion here to an earlier stage at which doubts about the Aristotelian view developed while still assuming that perception is passive.

Consider first the existence of perceptual illusions ranging from the octagonal tower that looks round when seen from a distance to the phantom limb. That illusions occur was not news in the seventeenth century, but they shifted from being considered unimportant curiosities to central phenomena that had to be dealt with by an adequate account of perceptual knowledge. This change of emphasis was supported by several overlapping strands in the developing science-philosophy complex, but I will consider only one strand here: the demand for a purely causal account of physical phenomena.

One direct consequence of a purely causal account of the physical world is that we must now think of sensory experience as initiated by the impact of physical objects on our sense organs and further mediated by causal processes in our nerves and brain. Once this step has been taken, two further reflections make it problematic whether physical objects ever have the properties that they appear to have when we perceive them. The first of these comes from reflecting on pain--a major example in the seventeenth century. An interaction between my body and another physical object may cause me to feel pain, but it is generally recognized that when we feel pain we are not perceiving a property that was out there in the physical object waiting to be detected under appropriate circumstances. Yet the causal process that yields our perception of any property of a physical object is of the same as that by which we come to feel pain.[2] Thus the experience of pain served as an entering wedge for arguing that sensory experience need not provide an accurate image of the physical object that our experience. But once we begin paying attention to examples of this sort, one standard problem of the epistemology of perception immediately arises: by what criteria can we assess which, if any, of the properties that a physical object appears to have actually characterizes that object.

Let me stress the key point that I am after in reviewing this bit of history: <u>the central epistemological question that I have just formulated arises because we believe that our senses operate in a particular way. Given a different account of the nature of our senses, or a radically different view</u>

of how we acquire information about the physical world, then this particular problem of epistemology need not arise.

This conclusion can be further illustrated by considering a second standard epistemological topic, the theory of meaning. Throughout the history of empiricism it has been maintained that meaningful language must somehow be tied to experience. In the twentieth century this doctrine has taken many forms. Among logical positivists it appeared as the thesis that a non-analytic proposition is meaningful if and only if there is a procedure by which that proposition can be empirically verified or falsified. Among later logical empiricists this became the claim that there is an observation language consisting of terms (not propositions) that get their meaning directly from experience; all other terms--including theoretical terms of science--must be either defined or somehow introduced on the basis of this observation language.[3] Other empiricists, who were not primarily concerned with philosophy of science, sought to reduce claims about material objects to claims about colors, shapes, feels, and other qualia that are immediately available to our senses. But why should we accept this dependence of meaning on experience? In the writings of the classical empiricists the answer was clear: Locke, Berkeley and Hume based their account of meaning on an empirical account of how people actually learn the meanings of words. Now many twentieth-century empiricists rejected this justification because they sought to establish a philosophical discipline of epistemology that is independent of empirical psychology. Yet it is difficult to find an explanation in the twentieth-century literature of why we should accept this account of meaning that does not require a foundation in psychology. The most detailed accounts consist of attempts to show that, with the aid of modern logic, we can define terms that refer to material objects on the basis of terms that refer only to sensible qualia. But even if the project succeeded--and it did not--we could still ask why we should proceed in this direction rather than attempting to define terms referring to sensible qualia on the basis of material object language. It would seem that the only plausible reason for preferring one direction to the other is some claim about human psychology. My point, again, is that the very way in which traditional epistemological problems have been formulated depends on empirical claims about human cognition.

3. The Anti-Naturalist Returns

I have been arguing that an attempt to construct a normative epistemology that is to be relevant to human beings must itself be built on an account of human cognitive abilities, but now it is time to give anti-naturalist epistemologists their due. For there remains a problem with our discussion thus far which will set the framework for the remainder of this paper. One task of normative epistemology is to establish principles for evaluating empirical claims against the available evidence. Now the body of psychology on which we must draw in order to arrive at an account of human cognition must be established empirically. But this requires that we already have a set of methodological rules that will allow us to evaluate empirical claims within psychology. We would appear to be caught in a vicious circle: we need a body of psychology in order to choose an appropriate methodology, but we need a methodology in order to evaluate psychological hypotheses. An example from recent psychology will underline the problem.

A number of psychological studies have led to the conclusion that human beings exhibit a confirmation bias: given a preferred hypothesis, they seek evidence that will support this hypothesis and tend to ignore evidence that contradicts the hypothesis (e.g., Nisbett and Ross, 1980; Faust 1984). Now consider two different ways of thinking about this result. We could accept confirmation bias as a universal feature of human cognition, take it as therefore normatively correct, and support the claim that confirmation bias is rampant by providing evidence that confirms this claim. If this seems unsatisfactory, we can note that the psychologists who report this result agree. They maintain that confirmation bias is normatively incorrect in spite

of its widespread occurrence. Indeed, they hold that there are systematic reasons why we should seek and give priority to evidence that contradicts our favored hypotheses. Yet in making this latter claim our psychologists are committed to the view that there are epistemic norms that can be justified on grounds other than empirical studies of human reasoning. Presumably, their own studies of human cognitive behavior adhere to these norms and do not exhibit confirmation bias. Thus it would appear that there must be some non-empirical justification for these norms and that this justification must be in place before we accept any of their claims about actual human reasoning.

We have now returned to the point from which this paper began and we can perhaps see more clearly why opponents of a naturalistic epistemology maintain that epistemic norms must be established a priori. A priori norms would provide an anchor outside of the process of empirical investigation that would provide a non-circular foundation for empirical inquiry. Yet the argument that took us from the beginning of this paper to the present point also reasserts itself. For once it is proposed that we need an a priori justification for our norms we must ask whether we human beings are capable of such a priori justifications and we are thrown right back into the circle since it is difficult to see how we can answer this question about human capabilities other than empirically.

So how shall we proceed? I will work from this point in two main stages. First, I shall leave the question of the foundations of methodology in the background and explore, on purely empirical grounds, some features of human cognitive abilities that are relevant for assessing our cognitive capabilities. And, on the basis of this discussion, I will propose some epistemic norms. Second, I will return to the foundational issues in section 10, examine the problem of vicious circularity more systematically than I have done so far, and then consider whether the arguments for the specific norms I have proposed are in fact vitiated by a circle.

4. The Naturalistic Perspective

The single most important development leading to naturalistic epistemologies is the theory of evolution with its inclusion of human beings in the evolutionary process. The key points I want to make have been argued elsewhere by many people (cf. Radnitzky and Bartley 1987; Halweg and Hooker 1989) and I will limit myself here to summarizing the relevant results.

Once we have learned to think of human beings as denizens of the natural world produced by the evolutionary process, we must give up any view of human cognition that assumes magical insights into reality. As a species, we began from ignorance and have acquired what knowledge we have through a slow, uncertain, non-linear process of trial and error with no guarantee of success. In one respect this thesis was already central to the empiricist tradition which held that all of our substantive knowledge of the world depends on experience and that any generalizations that go beyond what we have experienced are falsifiable and subject to replacement. But twentieth-century empiricists have been quite insistent that while this fallibilist view of cognition applies to our substantive knowledge of the world, it does not apply to our understanding of methodology. Methodology was conceived of as an a priori achievement along the lines developed above. However, a more thorough-going naturalism holds that our understanding of methodology is also arrived at by the kind of tentative process of theory construction that leads us to our understanding of various aspects of the world. This idea is captured in a slogan that has been adopted by several advocates of a thorough-going naturalism: we must learn how to learn about the world; what methods work depends on what the world is like and on what kind of being we are. (See, for example, Shapere 1984, Hooker 1987, Laudan 1987.) One consequence of this view is that our methodological and substantive research must interact because our methodological theory must at least be consistent with our best available understanding of the kind of cognitive equipment we have and the kind of world we are dealing with.

The basic epistemological lesson to be drawn from evolutionary theory, then, is that we should be rather modest about our knowledge-claims at every level--including methodology. And this lesson has been reinforced by developments in many fields. Gödel's incompleteness theorem is only one of a number of results that have led us to reexamine the epistemic status of mathematics, which has long stood as the pinnacle of certainty to which other subjects have aspired. (See Kline 1980, for many more examples.) Other limiting results, such as Church's theorem and results from complexity theory in computer science, have brought out limitations of a universal Turing machine, which has provided the model of an ideal computer and often a model of human cognition as well (e.g., Cherniak 1986; Reiner 1993). Meanwhile, studies in the psychology of human reasoning and the limitations of human memory have suggested that human cognition is significantly less powerful than the best that we can expect from a digital computer. (Cf. Wason and Johnson-Laird 1972; Nisbett and Ross 1980; Kahneman, Slovic and Tversky, 1982; Cherniak 1986, Ch. 3.)

Still, we must not be modest to the point of error. For while we have excellent reasons for relinquishing some of the more grandiose pretensions of our intellectual tradition, we should not ignore the rather impressive evidence suggesting that we have a much wider range of cognitive abilities than any other species we have encountered thus far--and that there may even be respects in which we have abilities that are not shared by a universal Turing Machine. Consider, here, our ability to develop complex systems of abstract mathematics, to design spacecraft to explore other planets even though this has no direct relevance to our survival and reproduction on this planet, our ability to develop and improve a theory of evolution, and our ability to study the nature of our own reasoning and to develop normative theories of logic and probability that allow us to recognize the limitations of our spontaneous reasoning. An adequate account of human cognitive abilities must be able to account for these accomplishments. All of which leads me to the next step in my argument: I want to discuss a cluster of human abilities that I think are central to our epistemic accomplishments but that have been given too little attention in the naturalistic literature. But I want to work my way towards this discussion by reviewing yet another theme from traditional epistemology.

5. The Logic of Justification

One of the pervasive concerns of traditional epistemology was generated by a plausible view of the logic of justification. On this view we justify a proposition, P, by properly relating it to other propositions. Let us use P* for the set of propositions invoked to justify P. Now for the proposed justification of P to succeed, each of the propositions constituting P* must itself be justified and, on the standard model of justification, the members of P* must be justified by their relation to some other set of propositions, P**. An infinite regress clearly threatens: any proposed justification requires some prior justification, which requires other prior justifications, ad infinitum. Traditionally epistemologists responded by seeking regress stoppers: some set of propositions that are justified without reference to any other propositions. These privileged propositions have been variously described as "self-evident," "self-justifying," and "indubitable." Once these privileged propositions have been established, they then provide the foundation for all further justifications.

There is a second regress that this view of justification faces, although it has received less attention in the literature. Justification requires rules that specify the relation between the propositions being justified and those that provide the justification. But not just any rules will do; a successful justification requires that we use the correct rules and thus our choice of rules must itself be justified, and we are on our way again. Here too it would seem that we require some self-justifying foundational rules to stop the regress. Elsewhere I have reviewed the relevant history and argued that attempts to provide these foundations have regularly amounted to the

claim that our cognitive abilities include a form of infallible intuition that, under appropriate circumstances, provides us with the required knowledge of the foundational propositions and rules. (See Brown 1988b Ch. 1 for a more detailed account of the path to a foundationalist epistemology and Ch. 2 for the historical review.)

Now this appeal to infallible intuitions has no place in a naturalistic epistemology. Moreover, we have excellent empirical grounds for concluding that our cognitive repertoire does not include any such ability. The historical landscape is littered with universal propositions that were put forward as self-evident but that turned out to be false. Contemporary studies of perception--in both psychology and philosophy--have cast serious doubt on the indubitability of our perception reports. And the history of logic provides us with a display of controversies and reconsiderations even within the realm of elementary deductive logic. (See Brown 1988b, sec. 2.7 for examples from logic.)

Several responses are available at this point. One possibility that some philosophers will always find attractive is to hew to the traditional demands for a legitimate justification, accept the conclusion that such justifications cannot be produced, and adopt some form of scepticism. Another possibility is to argue that since no indubitable refutation of the traditional approach has been provided, we should continue the search for self-evident foundations. Yet another approach is to take seriously the conclusion that we have no unchallengeable beliefs, apply this conclusion to the traditional framework itself, and look for other ways of understanding rational justification that are compatible with our current understanding of our cognitive abilities. This last approach opens up a number of possible paths and I will follow one of these here; this requires that we look again at our cognitive abilities.

6. Skills

There are two points from our discussion of justification that require clarification. First, my remarks about rules that relate propositions to other propositions and rules to other rules can be read in two different ways: as referring to objective relations among propositions or to processes taking place within the mind of a cognitive agent. I am concerned here with the latter situation--with what is required for a specific individual to arrive at a justified belief. From this perspective it is not enough that one proposition entail another proposition. If I am to use the first proposition as part of a justification for the second, I must recognize this entailment.

Second, we must distinguish two different senses in which a process may be rule-governed: a process may <u>conform</u> to a rule or it may proceed by <u>following</u> that rule. A falling stone conforms to rules but does not follow rules; cognitive behavior provides a paradigm of rule-following. Note one striking difference between the two cases: a cognitive agent can fail to follow a rule, either inadvertently or on purpose; the stone cannot fail to fall in accordance with its governing rules. Since our topic is cognition, I will only be concerned with processes that follow rules.

It is widely held that all cognition proceeds by following rules, but the traditional account of justification suggests that this need not be correct. For if we ask why intuition is capable of stopping an epistemic regress, one possible answer is that acts of intuition do not depend on rules. Now I noted above that <u>infallible</u> intuitions have no place in a naturalistic account of mind, but this leaves open the possibility that our cognitive repertoire includes process that have the following characteristics: they do not operate by following rules, they are fallible but yield results that are more reliable than chance, and they can be included in a naturalistic epistemology. Such processes would be sufficient to stop an epistemic regress--although in a fallible, and thus tentative, manner. I am going to argue that such processes do exist. I will refer to processes of this type as <u>judgments</u>, avoiding the term "intuition" because of its historical association with an infallible insight. Thus I am going to retain most of the traditional account of justification, eliminating only the demand that justifications must rest on a

foundation that is never subject to reconsideration.[4] My next task, then, is to explain the concept of judgment in more detail and to provide reasons for holding that humans have this ability. The ability to exercise judgment is an example of a cognitive skill that is continuous with physical skills. Examples of physical skills include the ability to ride a bicycle, catch and hit a baseball, or use a carpenter's tools. I will work my way towards an account of judgment by considering four characteristics of physical skills.[5]

(1) Skills are learned through demonstration and practice, rather than by learning and following a set of rules. Ability will typically improve with practice and deteriorate with lack of practice--although relearning will tend to be faster than initial learning. That we do not learn a skill just by memorizing a set of rules is particularly clear in the many cases in which no one is able to formulate a set of rules that will successfully guide this activity. And even in those cases in which we can formulate such rules, being able to state the rules will not be sufficient for carrying out the activity in a skillful manner. For example, a physicist who fully understands the mechanics of keeping a bicycle balanced, but who has never ridden, will still have to practice in order to learn to ride.

Although we do not learn skills by memorizing rules that we then follow, it could be argued that skill learning is a process of extracting rules in some non-conscious manner and then following those rules. But on reflection this is not an especially plausible hypothesis. Consider how the process of skill improvement would look on this account. One possibility is that the more skillful practitioner will have learned a larger set of rules than the less skillful person in order to deal with a variety of special circumstances. Yet if the improvement of a skill requires the acquisition of a growing set of ever more specific rules to cover special cases, we would expect performance to deteriorate as the number of rules to be searched and evaluated at each stage increases. This is exactly the opposite of what we find. Alternatively, it could be argued that in improving a skill we develop a better ability to apply the rules we have already learned. But how we are to understand this notion of "applying rules?" If this requires learning further rules for the application of rules we are back to the expanding set of rules just considered, and we may even be on the verge of another regress. If the proposal does not require learning new rules, but some other kind of process, then my main thesis has been accepted: learning and exercising a skill cannot be completely captured in following rules. Moreover, it now seems that even our ability to follow rules is based on a more fundamental ability that does not require that we follow rules.

(2) For most physical skills, and especially for the more complex skills, there is a substantial range in which most people can learn and improve. Still, some people will learn much more quickly than others; some will develop a skill to a higher degree than others; and some will be unable to learn a particular skill at all. For example, most people can develop some skill at catching fly balls, but few achieve the level of a major-league ball player, and a few will be unable to develop this skill at all.

(3) The ability to exercise a physical skill is fallible in the sense that even the best practitioner will not always exercise that skill to the highest degree that is within her capability and in some cases may fail altogether. Yet an enormous variety of human accomplishments depends on physical skills. Note especially that the fact that even the most adept sometimes fail does not provide grounds for concluding that there is no such thing as a skill, nor any reason for refusing to invest greater confidence in those who have shown facility at a particular task than in those who have not.

(4) The above discussion leads directly to the central feature that concerns me: the ability to exercise any skill at all requires that we have some basic abilities that are not themselves carried out by following rules.

I want to spend a moment on one more attempt to account for skillful behavior wholly on the basis of following rules. In many cases we can examine a form of behavior, formulate a set of rules for that behavior, and then use

these rules as the basis for designing machines that will carry out this behavior. Often the machine will be more reliable than would a human actor. Moreover, this occurs not only in cases such as the construction of industrial robots, but also when we program a computer to carry out a task. There are, however, several points about this possibility that should be kept in mind if we are to avoid misunderstandings. First, the fact that we can study a form of human behavior and capture it in a set of rules does not show that the humans we have studied were following those rules. We engage in the same kind of analysis for falling stones, the motions of the planets, and myriad other phenomena that provide paradigm cases in which skill is not involved. Second, an industrial robot no more follows rules than does a falling stone. We may have designed the robot to act in accordance with a set of rules, but the robot's own behavior is strictly determined in the same sense that the velocity with which a stone hits the ground is determined once we have determined the height from which it is dropped. Third, the points just made about the robot apply equally to the digital computer. Whatever is involved in writing a program and entering that program into the computer, it is a mistake to think that a set of rules now exist in the computer and that the computer follows those rule. Entering a program amounts to changing the states of a number of circuit elements and thus creating an initial condition in the computer. The situation is analogous to that in which I might "program" a rock to hit the floor with a velocity of (roughly) sixteen feet per second by lifting it to a height of four feet and then dropping it.

Now let us consider cognitive skills. Paradigm examples include the ability to solve deduction problems in logic, play chess, invent algorithms, formulate the rules in accordance with which various process take place, and invent new scientific theories. These abilities clearly share the first four characteristics of physical skills noted above: they are taught by demonstration and learned through practice; there is a range of talents in developing and exercising these skills; and the exercise of these skills is fallible. The remaining feature--that cognitive skills are not exercised by following rules--provides the crux of the issue before us.

Before we tackle this issue, a distinction is in order. Many activities are defined by an explicit set of rules, but once we enter into this activity, there are different degrees of skill among those who obey these defining rules. All chess players have learned the rules that define the game, but knowing these rules does not automatically generate skillful play. In a similar way, those who write computer programs are constrained by the syntax of their programming language, but these rules do not generate algorithms--let alone efficient algorithms. The question currently before us is whether skillful performance results from following rules.

There are, of course, some rules for generating a skillful performance: rules for good chess play, programming heuristics, and so forth. But, again, these rules are not sufficient to generate a skillful performance in these areas. Not only do we find considerable variations in skill among those who take guidance from these rules, but it is just at the point where available rules cease to be sufficient that differences of skill become apparent. The construction of proofs in deductive logic will provide a useful example. If we are working within a specific deductive system every permissible step is governed by a finite set of rules, but there is no sufficient set of rules for deciding what step should be made at any given juncture. Moreover, we can easily formulate meta-rules for constructing proofs that are completely in accord with the laws of logic but that it would be utterly absurd to follow. Consider a typical textbook problem: we have a set of premises from which we must derive a specified conclusion. Suppose that two sentences, "p" and "q", are either among the premises or have already been derived, and consider the following procedure. First we use the rule of conjunction to derive the single sentence "p&q." Next we use simplification to deduce "p" again and "q" again. Then we apply conjunction once again, and so on until we drop. This would be a mindless application of a set of rules, but note that no law of

logic is ever violated in the procedure. The example is enough to make the point that following an established set of rules--even elementary rules of logic--is not sufficient to assure that we are behaving in a skillful manner.

A more crucial question is whether following a set of rules is necessary for skillful behavior, but at this point the traditional regress argument comes into play. Consider our example from logic once again. Someone who is skilled at constructing deductive proofs will be able to make a sensible choice of which rules to apply at each stage. We can look here for some meta-rules, but not just any meta-rules will do. Here too we can formulate lots of silly rules for making the decision. So we face another choice among rules and the need for even higher-level rules, and so forth. At some point the process must stop if we are ever to go forward and this stopping point cannot itself be chosen by yet another set of rules. In other words, if justified rules are necessary for every coherent cognitive choice, then no choices will be possible. Yet we do produce deductive proofs and my proposal is that the ability to end the regress--for example by choosing the next step in the proof--is the exercise of a skill. Not every choice will be correct--skills are fallible--but the choices of a skillful logician will be more reliable than chance. In other words, skills are regress-stoppers, but tentative, fallible regress-stoppers.

I want to make one more point before applying this account to judgment. The argument thus far can be seen as attempting to establish a close analogy between physical and cognitive skills, but an evolutionary perspective suggests that more than an analogy is involved. If a human being is a complex but unified organism, we should find genuine continuity between physical and cognitive skills. And there are cases--perhaps best illustrated by activities such as painting, sculpture and musical improvisation--in which the physical and cognitive aspects of a performance are so thoroughly integrated as to defy any clear separation.

7. Judgment

Judgment is a generic notion that refers to a class of cognitive skills: the ability to make decisions in a specific domain without relying on a set of rules. We must rely on judgment in many situations, for example when no appropriate rules are known, or when there are alternative rules among which we must choose, or when we recognize that established rules cannot be used within available time limits. Several points about the notion of judgment must be stressed. First, judgment is domain specific: one develops engineering judgment or chess judgment or logical judgment much as one develops specific physical skills--by studying available knowledge (including established rules and techniques) in a particular field and practicing in that field. Second, judgment is exercised with respect to specific problems and decisions and thus requires study of the available information relevant to the problem at hand. A chess master does not recommend a move without studying the game in question, or a physician recommend a treatment without studying the patient. Third, it follows from the first two points that not every choice made without following rules is an exercise of judgment. Judgment requires specific expertise and information and there is no more guarantee that we will actually be able to exercise judgment in a given case than that we will know a rule. Fourth, while the exercise of judgment is an individual ability, it is not "subjective" in any pejorative sense. Judgment is exercised in response to specific problems and in the light of information that typically comes from outside of the individual. In other words, the ability to exercise judgment is no more "subjective" than is the ability to drive a car or catch a fly ball.[6]

We must now ask whether judgment has a place in a naturalistic account of human beings. Even though judgment lacks the infallibility that was traditionally attributed to intuitions, what grounds are there for believing that natural systems can embody such abilities? Within the limits of this paper I can only point to one recent development that provides strong reasons for holding that physical systems are capable of such behavior. I refer here

to neural net computers that can make distinctions and arrive at reliable results without having been programmed in accordance with a set of rules or built so as to embody some set of rules. These computers are "trained" to carry out a specific task in much the way that human beings are trained. The trainer begin with a set of inputs for which the correct outputs are known and the machine's internal structure is adjusted slightly when it gives a wrong result. The machine's accuracy improves over time as the training process continues and in the course of training the machine will often make mistakes of the same kind that people make as they learn a particular skill. Indeed, not only is a program never written to govern the machine's behavior, but in some cases the trainers do not even understand how the machine has arrived at its successful behavior.[7]

Finally, the claim that humans can develop and exercise judgment is an empirical thesis. I have offered some anecdotal evidence for this claim, but the claim should be susceptible to systematic empirical study; systematic evidence that the phenomenon does not exist will undermine my proposal.

8. From Judgment to Reason

Let us now consider how judgment is to be incorporated into a full-blown account of rationality. The key point is that while rationality must ultimately rely on judgment, there are means available for improving the reliability of our judgments. Full rationality requires making use of these means in order to move from an initial judgment to a more reliable, better founded judgment. All of these means of improving judgment share the following feature: they increase the objectivity of judgment by taking the individual outside the circle of her own reflections. There appear to be three major ways of pursuing this improvement of judgment: the use of formal methods, observation, and submission of one's results to others for critical debate. Only brief comment on each these will be possible in the present paper; in each case I want to illustrate how we can impose additional constraints on individual judgments.[8]

Formal methods draw out consequences of our claims that are implicit in those claims but that are not obvious. For example, the result of a series of arithmetic operations on a specific set of numbers is already determined; doing the arithmetic allows us to discover this result. In a similar way, whether a set of propositions is consistent is a fact about that set--a fact about which we may be mistaken. The classic example is Cantor's set theory which was assumed to be consistent until Russell's proved otherwise.

Observation is not relevant to all cognitive tasks but where it is relevant it provides constraints on our theorizing by bringing us into contact with the items about which we are making claims and allowing those items to have their say. Note that I am <u>not</u> maintaining that observational claims are <u>ipso facto</u> more certain than other kinds of claims. Rather, observation provides special constraints on our beliefs because observation brings us into contact with a world that is independent of those beliefs.

Finally, we should submit our views to evaluation and criticism by others because other people will have perspectives, information and skills that we do not have. To be sure, decisions as to how much observational data we should collect and when to cut off the processes of inference and debate are ultimately dependent on individual judgments, but there is nothing pernicious in this as long as we are clear that having arrived at a rational conclusion does not guarantee that the conclusion is true or that we have achieved the last word on the subject. Reason is a means of arriving at reliable but fallible results, results that can be reexamined when reasons for doing so appear.

9. From Descriptions to Norms

In the course of this discussion we appear to have moved quite naturally from descriptive claims about human cognitive abilities and about means that are available for increasing their reliability to a number of norms for rational evaluation of our views. Here are some of these norms:

N1. Do not attempt to exercise judgment on matters in which you lack appropriate expertise.
N2. Even where you have expertise, study the details of the problem at hand before arriving at a judgment.
N3. Always consider your judgments to be fallible and subject to further evaluation.
N4. When evaluating the results of a judgment pay attention to the logical consequences of that result.
N5. Where appropriate, gather observational evidence and seek to extend the range of available evidence.
N6. Submit your results to others for suggestions and criticism.

Although these norms have certainly not been deduced from the preceding descriptions, there is still a sense in which these norms have been based on those descriptions. Assuming that my descriptive claims about human cognitive abilities are correct, I have put forward a number of hypotheses about how to pursue reliable beliefs given these abilities. These norms are hypotheses in the sense that they too are subject to evaluation. One point of evaluation has already been noted: if empirical studies were to show that human beings have no ability to exercise judgment that is more reliable than chance, then the motivation and a major part of the justification for these norms will have been removed. In a similar way, if one could show that submitting one's views for critical debate tends to confuse issues and reduce the reliability of our beliefs, then N6 would be brought into serious question.

Note also that I am pursuing these norms in a rational manner in exactly the sense that I seek to capture in these norms. I am working in an area in which I believe that I have relevant expertise, that view has received some support from others in the field, I have studied a significant portion of the relevant literature, and by publishing my ideas on the subject I am submitting them to others for critical debate. Thus my project exhibits an important kind of reflexivity and it would seem that this reflexivity is unavoidable. After all, I am attempting to provide an account of rationality and I wish to pursue this attempt in a rational manner. Therefore I will certainly work in terms of the account of rationality that I think to be correct. But at this point the traditional worries about a vicious circle reappear. I cannot find any vicious circularity in the procedure I have followed, but the charge of circularity is a general one and, if it could be sustained, then my failure to find a vicious circle in my own procedure would provide no more (an no less) than a comment on the limits of my vision. It is important then that we attack the charge of circularity head-on.

10. The Circle Argument

The major objection to using scientific results in order to justify the scientific enterprise is that any such attempt begs the question--it assumes the very conclusion that it seeks to establish (e.g., Stroud 1985, Siegel 1989). But while this may seem a devastating objection, I shall argue that it is far from conclusive. Some preliminary points will set the stage.

First, there are areas in which some form of circularity seems to be unavoidable. For example, an empirical study of our perceptual abilities will presumably make use of those abilities. Similarly, attempts to prove that the system of logic we currently accept is correct will require that we make use of some logic in the proof. Presumably we will use that system of logic that we believe to be correct, rather than some alternative system that we consider to be in error.

Second, there are situations in which self-reference is not only unavoidable, but mandatory. For example, if my theory of knowledge holds that human beings are only capable of fallible conjectures, then I must treat that theory as a fallible conjecture on pain of self-refutation. In the same way, a theory of our cognitive capacities that is so restrictive as to entail that it is impossible for us to have produced and defended that very theory is ipso facto unacceptable.[9] This is why charges of reflexive inconsistency are so devastating to the more extreme forms of relativism: their proponents appear

to claim that their view is correct in a sense to which they deny any significance and on grounds that they maintain are not determinative of rational belief.

Third, it is an odd feature of question-begging arguments that, as a matter of logical form alone, they are valid. Whatever is wrong with question-begging arguments, the problem must be located in something other than their logical form. Let us consider, then, just what is problematic about the class of circular arguments that concern us here--those that assume a theory or a methodology as part of an attempt to justify that very theory or methodology. We can to approach this question by returning to the phenomenon of confirmation bias and considering why this tendency should be avoided. The key point has been familiar since Popper: it is always possible to find some confirming evidence for any hypothesis. But even in the face of massive confirming evidence, our hypothesis may still be false. On the other hand, if a false conclusion has been deduced from a set of premises, then something is certainly wrong with those premises; and if a set of propositions is inconsistent, it is not possible that all of the propositions in the set are true. This asymmetry between confirmation and disconfirmation suggests a basic principle of methodology: a proposed test of an hypothesis is legitimate only if it possible that the test result in rejection of that hypothesis.[10]

Now those who argue that a naturalistic justification of methodology will be viciously circular are worried about a particularly drastic kind of confirmation bias: they are concerned that a test of an hypothesis, H, that assumes H in an essential way could not possibly refute H. Any such test, it is maintained, would necessarily come out in favor of H, but this support would be illusory because H could not possibly have been challenged by the test. Some contemporary skeptics about reason have made exactly this point:

> The basic point is that justifications of deduction themselves presuppose deduction. They are circular because they appeal to the very principles of inference that are in question. In this respect the justification of deduction is in the same predicament as the justifications of induction which tacitly makes inductive moves by appealing to the fact that induction 'works'. Our two basic modes of reasoning are in an equally hopeless state with regard to their rational justification. (Barnes and Bloor 1982, pp. 40-41)

This is a seductive thesis but, I shall argue, it is no more than that. It seems plausible as long as we keep the discussion on a very high level of abstraction, but when we look at specific examples in sufficient detail, we will see that there are in fact cases in which assuming the hypothesis being evaluated does not prevent us from arriving at a challenge to that hypothesis. I will develop this point here for a particularly important example, the case of elementary deductive logic.[11]

Consider a system of logic, L_1, which allows the argument form A: from a premise of the form "p or q" we may deduce "p". By definition, an argument form is invalid if there is an argument of that form which has all true premises and a false conclusion. Consider, then, the following argument in which A itself plays a central role. Let "p" be "A is invalid," which we are assuming to be false; let "q" be any true proposition that you like. Our argument is:

> A is invalid or q.
> Thus, A is invalid.

The premise is true and the conclusion is, by hypothesis, false. Thus the argument form we are examining is invalid. However one may respond to this result, one point is clear: the fact that our argument proceeded on the assumption that A is valid did not prevent the argument from yielding a challenge to the validity of A.

As a second example consider a system of logic, L_2, that is inconsistent although those who are using this logic do not know this. Suppose we set out to study the consistency of L_2 and, since we believe that L_2 is correct, we

carry out our study in a metalanguage that uses L_2 as its logic. Given that any proposition can be validly deduced using an inconsistent logic, then it is certainly possible for us to arrive the conclusion that L_2 is inconsistent.

The crucial point in both of these examples is that assuming a particular system of logic does not prevent us from deducing a conclusion that challenges the correctness of that very system. A moment's reflection will suggest that neither of the examples we have just considered should be especially surprising. Logics that are internally inconsistent or that include invalid arguments share the feature that they are not sufficiently restrictive. Given their excess liberality in licensing inferences, they actually make it easier to construct self-refuting arguments than would be the case for a more restrictive logic.

Let me underline the moral of these examples. They show that a circular argument does not automatically embody a vicious circle. As a result, those who wish to reject an argument because it "assumes what it seeks to prove" must themselves prove that the specific circle in question is vicious. In the particular case of my attempt to offer an account of normative rationality that is normatively rational on its own grounds, I will rest with the point that I can find no vitiating circularity until someone points one out.

11. Conclusion

A naturalistic epistemology should seem attractive--even natural--to those of us who view humans as members of the natural world. It is, thus, an odd feature of contemporary philosophy that many philosophers who reject any appeal to the supernatural in giving an account of human beings also reject the attempt to construct a naturalistic account of scientific knowledge. The objections, we have seen, derive from two closely related concerns: the worry that science cannot provide epistemic norms since science is a descriptive endeavor and we cannot deduce norms from descriptions; and the claim that even if we could extract norms from science, the attempt to use these norms to justify scientific results would be viciously circular. I have responded to both of these concerns in the present paper. With respect to the first I have argued that deduction is not at issue in arriving at norms for knowledge. Rather, we can reflect on our best understanding of how human cognition operates and on our cognitive goals and make proposal as to how best to pursue those goals. Moreover, these proposals can be critically evaluated. There is more to critical evaluation than has been examined in this paper (cf. Brown 1988b), but we have seen that one major form of evaluation derives from the account of human cognition that we accept. Epistemic norms that are based on a particular account of our cognitive abilities become suspect if that account is rejected, and norms that require us to do what is beyond our capabilities are surely unacceptable.

My response to the second concern was to argue that, once we look at details, rather than at general pronouncements, we find that reflexive evaluations need not involve a problematic circularity. Claims of vicious circularity must be justified by demonstrating that the specific case in question does embody a vicious circle.

The main outcome of both lines of argument is to focus philosophical attention on the details. Whether a naturalistic account of reason is possible depends on what cognitive abilities can appear in the natural world and this question cannot be answered a priori. Rather, as we develop a better understanding of actual cognition we can also develop a better grasp of the appropriate norms for us and we can apply these norms to provide a more accurate evaluation of our cognitive theories. In metaphorical terms, we have a helix, not a circle, and the helix is anchored in studies of actual human knowers. If there is something logically defective about this process, it will have to be show by examining specific cases.

1. Let me remind the reader that philosophers and logicians use the term "deduce" in a much more precise and narrow sense than it is used in much common discourse. Sherlock Holmes, Mr. Spock, and most people most of the time use "deduce" as a synonym for "infer" and include everything from rigorous deductions to educated guesses in its scope. The narrow sense of the terms limits deduction to inferences in which the truth of the conclusion is guaranteed by the truth of the premises. I shall use "deduce" only in this narrower sense.

2. Consider: I need sufficient light to see the color of the wall; I must rub my hand across the wall with a moderate degree of pressure to feel its inherent roughness; I must press fairly hard to feel its inherent solidity; I must slam my hand on the wall really hard to feel the wall's inherent pain.

3. See Hempel 1965 and Brown 1979, Ch. 3 for discussion of the changing accounts of how non-observation terms are to be introduced.

4. I want to emphasize that such processes are <u>sufficient</u> to end the regress. Other hypotheses as to how this occurs, and more radical departures from the traditional account of justification, are surely possible.

5. A sketch of a similar view will be found in Suppes 1984, pp. 215-219; see also Brown 1988, Ch. 4. Dreyfus and Dreyfus 1986, 1991, have proposed a detailed account of how skills are learned.

6. The ability to exercise judgment is typically learned in a social setting, although the early stages in the development of a new kind of skill may provide an exception. But this does not alter the point that each exercise of judgment is an individual act.

7. See, for example, Rumelhart, McClelland, <u>et al</u>. 1986 and Churchland 1989. See also Brown 1988, sec 4.3 for further discussion of the place of judgment in the natural world.

8. I have discussed the role of observation in Brown 1987 and 1990, and of critical debate in Brown 1988, ch. 5. All three will be fully integrated in Brown and Hooker, in preparation.

9. Siegel maintains that epistemological theories must be self-justifying (1989, p. 371).

10. This is another methodological rule that is suggested, but not entailed, by a set of facts--in this case facts about logical relationships.

11. See Brown forthcoming for an example that makes the analogous point in the case of observational tests of a theory.

References

Barnes, B. and Bloor, D. (1982), "Relativism, Rationalism and the Sociology of Science," in M. Hollis and S. Lukes, eds., Rationality and Relativism. Cambridge: MIT Press.

Brown, H. I. (1979), Perception, Theory and Commitment: The New Philosophy of Science. Chicago: University of Chicago Press.

Brown, H. I. (1987), Observation and Objectivity. New York: Oxford University Press.

Brown, H. I. (1988a), "Normative Epistemology and Naturalized Epistemology," Inquiry 31, pp. 53-78.

Brown, H. I. (1988b), Rationality. London: Routledge.

Brown, H. I. (1990), "Prospective Realism," Studies in History and Philosophy of Science 21, pp. 211-42.

Brown, H. I. (forthcoming), "A Theory-Laden Observation Can Test the Theory," British Journal for the Philosophy of Science.

Brown, H. I. and Hooker, C. A., Beyond Formal Reason, in preparation.

Cherniak, C. (1986), Minimal Rationality. Cambridge: MIT Press, 1986.

Churchland, P. M. (1989), The Neurocomputational Perspective: Philosophical Essays on the Mind, Brain, and Science. Cambridge: MIT Press.

Dreyfus, H. and Dreyfus S. (1986), Mind Over Machine. New York: The Free Press.

Dreyfus, H. and Dreyfus S. (1991), "Towards a Phenomenology of Ethical Expertise," Human Studies 14, pp. 229-250.

Faust, D. (1984), The Limits of Scientific Reasoning. Minneapolis: University of Minnesota Press.

Halweg, K. and Hooker, C. A. (1989), Issues in Evolutionary Epistemology. Albany: SUNY Press.

Hempel, C. (1965), "Empiricist Criteria of Cognitive Significance: Problems and Changes," in Aspects of Scientific Explanation. New York: Free Press.

Hooker, C. A. (1987), A Realistic Theory of Science. Albany: SUNY Press.

Hume, D. (1978), A Treatise of Human Nature 2nd edition, L. A. Selby-Bigge, ed., revised by P. H. Nidditch. Oxford: Oxford University Press.

Kahneman, D., Slovic, P. and Tversky, A. (1982), Judgement Under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press.

Kline, M. (1980), Mathematics: The Loss of Certainty. New York: Oxford University Press.

Laudan, L. (1987), "Progress or Rationality? The Prospects for Normative Naturalism," American Philosophical Quarterly 24, pp. 19-31.

Lindberg, D. (1976), Theories of Vision from Al-Kindi to Kepler. Chicago: University of Chicago Press.

Nisbett, R. and Ross, L. (1980), Human Inference: Strategies and Shortcomings of Social Judgment. Englewood Cliffs, NJ: Prentice-Hall.

Radnitzky G., and Bartley, W. W. (1987), Evolutionary Epistemology, Rationality, and the Sociology of Knowledge. La Salle, IL: Open Court.

Reiner, R. (1993), A Framework for Theories of Bounded Rationality. Unpublished PhD dissertation, York University.

Rumelhart, D., McClelland, J. et al. (1986), Parallel Distributed Processing, 2 vols. Cambridge: MIT Press.

Shapere, D. (1984), Reason and the Search for Knowledge. Dordrecht: D. Reidel.

Siegel, H. (1989), "Philosophy of Science Naturalized? Some Problems with Giere's Naturalism," Studies in History and Philosophy of Science 20, pp. 365-375.

Stroud, B. (1985), "The Significance of Naturalized Epistemology," in H. Kornblith, ed., Naturalizing Epistemology. Cambridge: MIT Press.

Suppes, P. (1984), Probabilistic Metaphysics. Oxford: Oxford University Press.

Wason, P. and Johnson-Laird, W. (1972), Psychology of Reasoning. Cambridge: Harvard University Press.