A Naturalistic Perspective on Epistemic Norms

Louise Antony
The Ohio State University

Naturalized epistemology, as I understand it, is the practice of treating knowledge—human or otherwise—as a natural phenomenon, susceptible of investigation by the methods of empirical science. A naturalized approach to the study of knowledge differs saliently from more traditional forms of epistemology in taking the existence of knowledge for granted. Naturalized epistemologists do not concern themselves with skeptical challenges. Nor are naturalized epistemologists much concerned with questions about what counts as “knowledge,” properly speaking. They do not worry if a bird’s natively specified program for star-based navigation is “justified” for the bird, nor if the sub-personal data structures and algorithms posited by cognitive psychologists can be properly counted as “beliefs.” The naturalized epistemologist is interested in the explanation of anything that even appears to be a cognitive achievement, whether or not it passes muster as “knowledge” in some preferred sense.

This indifference to what some philosophers regard as the defining issues of epistemology has provoked the charge that naturalized epistemology is not really epistemology at all. According to these critics, the problem of skepticism and the analysis of knowledge are part and parcel of epistemology’s normative charge: to specify the conditions for good knowing. But, they claim, naturalized epistemology offers only descriptions of cognitive processes, replacing normative accounts of the relation between evidence and theory with genetic, causal accounts. As Kim puts the complaint, “Epistemology is to go out of the business of
In this paper, I want to respond to these critics. Not only does a naturalized epistemology have a place for epistemic norms, it provides a more satisfying account of the grounding of those norms than do more traditional epistemologies. A naturalistic approach to knowledge can tell us not only what we ought to do, epistemically speaking, but why we ought to do it. Furthermore, the norms delivered so delivered will be useful norms – they will be tailored to the demands made by the real world on creatures embodied in the way we are embodied.

The paper has three parts: in Part I, I’ll try to answer the critics who think that a fully naturalized epistemology leaves no room for the normative. In Part II, I’ll explain what I think a naturalized approach has to tell us about the epistemic norm of objectivity, construed as strict impartiality. Here I’ll draw out a parallel with certain criticisms, within moral philosophy, of the ethical norm of impartiality. In both the epistemic and the ethical cases, I’ll argue, the naturalistic perspective counsels us to attend to the actual conditions under which human beings do things -- how they seek to know and how they strive to act rightly. In Part III, I will draw some conclusions about the function of norms.

I

Naturalized epistemology is often characterized in terms of what Hilary Kornblith has called “the replacement thesis,” viz., the thesis that traditional epistemology ought to be abandoned in favor of empirical psychology (Kornblith 1994a, 4). It is this radical thesis that most disturbs critics of naturalized epistemology, for it seems to them to either conflate the question of how we ought to arrive at our beliefs with the question of how we do arrive at our beliefs, or else to ignore the normative question altogether. Importantly, naturalized epistemology thus conceived is held to entail not only a flat-out rejection of the normative aims
of traditional epistemology (Kim in Kornblith 1995, 40), but also the repudiation of all \textit{a priori} elements in the study of knowledge (Haack, 1993, 119-20). I’ll return shortly to this presumed connection between the \textit{a priori} and the normative.

Indeed, very few contemporary epistemologists are willing to defend naturalism in this form: even those who identify themselves as proponents of a naturalized approach to knowledge hasten to qualify their commitment. These epistemologists prefer to defend some weaker form of naturalism – one that holds that traditional epistemology is not to be replaced, but only to be somehow constrained or informed by the empirical study of epistemic processes. Thus Kornblith offers for our consideration a weakened version of the replacement thesis: “psychology and epistemology provide two different avenues for arriving at the same place.” (Kornblith, 1994a, 7) If this thesis, rather than the strong replacement thesis is true, then, Kornblith tells us, there is no danger of traditional epistemology’s being replaced by or eliminated in favor of a wholly descriptive, wholly a posteriori science: “If the [weak replacement] thesis is true, the psychology of belief acquisition and epistemology are two different fields, which ask different but equally legitimate questions and have different methodologies.” Susan Haack says that she wants to defend a “modestly naturalistic position” according to which traditional apriorism will take its place as “the philosophical component of a joint enterprise with the sciences of cognition” (Haack 1993, 118).

Of course these “weak” and “modest” forms of naturalized epistemology have all the defects of liberal compromise over radical clarity: they forge a false consensus by making the proposal so vague that no one can disagree with it, meanwhile doing very little to alter the status quo. What’s the alternative to “naturalism,” thus construed? The view that empirical psychology has \textit{nothing} to offer the epistemologist? Who wants to be stuck saying that?
Nobody wants to be naively aprioristic, any more than anyone wants to be crudely scientistic.¹

There’s a more forthright way, I think, to answer the concerns about the elimination of norms within a naturalized framework, and that is to challenge the conception of empirical enquiry presupposed by those who are scandalsized by the “strong” version of naturalized epistemology. There are points in particular that I dispute. First, the critics must be presuming that science is a purely descriptive enterprise – that it neither contains nor can yield information about norms; why else think that a strongly naturalized epistemology can contain no normativity? Second, there is the presumption that science, and hence, a strongly naturalized epistemology, can make no appeal to, nor give any account of, the a priori. Perhaps this seems obvious. If epistemology is to be replaced by empirical inquiry, what role is left for a priori knowledge? If even logic is subject to empirical disconfirmation, what could possibly be known a priori?

I do not want to enter into an exegetical debate about Quine (although I may not be able to resist citing passages that support my interpretation). Let me just say that there is a reading of Quine according to which he was not rejecting the idea of a normative epistemology tout court, but was rather criticizing one particular approach to normative epistemology. As I interpret Quine’s critique of positivism, the problem he saw was not simply that positivistic epistemology relied on an indefensible distinction between the analytic and the synthetic. Rather, I think Quine saw positivism as failing by its own lights in the project of vindicating scientific practice. Carnap (at least on Quine’s reading of him) offered “rational reconstruction” as a solution to a

¹ See, for example, Susan Haack: “...mine is, in a sense, a naturalistic epistemology: it is not wholly a priori, since it relies on empirical assumptions about human beings’ cognitive capacities and limitations, and so acknowledges the contributory relevance to epistemology of natural-scientific studies of cognition. But this modest naturalism is very different from the much more radical, scientistic approaches which also go by the title, “naturalistic epistemology.” (Haack 1993, 4).
problem in *normative epistemology*, one that arises on an empiricistic view of knowledge: viz., if it is experience and experience alone that justifies belief, how can we justify belief in a science that contains both a *priori* elements and references to unobservable entities and processes? Carnap’s solution was to segregate the *a priori* elements of theories from the empirical elements, explain the *a priori* elements as “conventional” and then c) display the empirical elements as arranged in a hierarchy of dependence on sensory data, where the structure of the hierarchy is determined by the conventions that generate the *a priori* elements.

Quine did, of course, object to every element of this proposal. But the overall thrust of the critique was global: Carnap’s characterization of the structure of scientific confirmation was *false*, and a false story can vindicate nothing. If what we want to understand is how empirical knowledge is possible, we gain nothing by telling just-so stories about how theories *would* have been justified *had* they been derived in a certain way from the evidence.

But why all this creative reconstruction, all this make-believe? 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 proceeds? Why not settle for psychology?²

The idealizing inherent in rational reconstruction strips away exactly the features of actual enquiry that make the matter of vindication pressing in the first place – better to confront squarely the fact that we seem somehow, by processes that do not match those that seem, *a priori*, necessary for epistemic reliability, to obtain substantive bodies of knowledge about the external world.

---

Construed in this way, a naturalized approach to knowledge points toward a strongly externalist, reliabilist epistemology. The existence of knowledge is the starting point, the explanandum, and the question of warrant becomes the question of what processes and procedures do, as a matter of empirical fact, reliably yield information. If a process works, it works – there is only the question of understanding how it works. If sensory data vastly underdetermines theory, and yet theory arises, and is tested, and continues to survive, it must be that something is being contributed by the subject, and that those contributions are salutary. As Quine himself emphasized, in The Web of Belief, “Natural Kinds,” and elsewhere, such “extra-empirical” features of theorizing as our preference for simplicitity and our prior judgments of similarity appear to be, for us, pragmatically necessary a priori impositions on hypotheses – impositions that can nonetheless be vindicated a posteriori by the overall success of the theories that result.

The starting point of such a naturalized, reliabilist epistemology must be what we know – or can come to discover – about the knowing subject herself. Although Quine’s commitment to behavioristic psychology – motivated, ironically, by philosophical arguments against the legitimacy of mentalistic posits – prevented him from embracing cognitivist psychology, it would appear that Noam Chomsky’s work on language acquisition constitutes the parade case for the naturalized approach. Begin with a mundane epistemic achievement: human children’s acquisition, within the first few years of life, of an intricate symbolic system that can be used for the expression and communication of messages unbounded in both their variety and their complexity. Observe the actual conditions under which this achievement occurs: exposure to a body of data that is both flawed and fragmentary, at least relative to the demands of theories of general learning. (This remains true whether we construe “learning” mentalistically, in terms of
the operation of a general-purpose inductive logic, or behavioristically, in terms of patterns of conditioned responses to environmental stimuli. Things look, indeed, somewhat worse on the latter picture.) Conclusion: the children are getting unauthorized help – God is telling them the answers.

The rehabilitation of mentalistic psychology, not to mention nativism, is therefore, on my view, a natural consequence of putting into practice Quine’s admonitions to eschew armchair philosophizing about the nature of knowledge. But once we recognize that there is an independent variable in the construction of knowledge inside the subject’s head, we have a new basis upon which the traditional distinction between the *a priori* and the *a posteriori* can be – partially, but satisfactorily – reconstructed. The distinction now becomes primarily an *etiological* distinction – an “inside/outside” distinction. The *a priori* elements of knowledge are those elements that reflect or derive solely from features of our cognitive machinery; knowledge that additionally involves information supplied by the senses is *a posteriori*.

Consider, for example, the fact that I am inclined to infer (3) “It’s Belgium” from (1) “It’s Friday” and (2) “If it’s Friday, then this is Belgium.” That (3) follows from (1) and (2) is something that everyone would say I know *a priori* (if they would say that I know anything *a priori*). On my account, that’s just to say that my inclination to perform the inference reflects the design of my cognitive machinery. I have, I hypothesize, a syntactic engine inside my head, and it is in virtue of its operation that I conclude (3) on the basis of (1) and (2).

Now maybe someone will raise the objection – I hope so, because I have an answer to it – that explaining *what causes me* to infer (3) from (1) and (2) is one thing, and explaining how I am *warranted* in so inferring is quite another. Indeed. It hardly follows from the mere fact that I am built to reason in a certain way, or that I am built to believe certain propositions, for that
matter, that I have *good epistemic reason* to do so, that I *ought* to so reason or to so believe. Innateness and apriority are not the same thing. There is always, in addition to the “genetic” question, the question of warrant or justification.

Notice, first, that nothing in my naturalistic programme has forestalled the *normative* question just asked; nothing has prevented the question of my *justification* for indulging a certain epistemic habit from arising. But secondly, notice that my reliabilism tells us exactly where to look for an answer. Look to the processes that produce the belief; look to the mechanisms that subserve the inference. Are they, or are they not, *reliable*? If they are, then I have all the warrant I can ever need (or expect to get). If I have reliable cognitive machinery, then my reasoning will tend to track the truth.

Ah, but how do I *know* that my cognitive machinery tracks the truth? As a reliabilist, I am within my rights to refuse to answer this question, because it has nothing to do with the question of warrant for my ground-level inferences. I am warranted in performing the inferences I perform just in case the subserving mechanisms are reliable, and it can hardly be maintained that I need to *know* them to be reliable in order for them to *be* reliable. To insist otherwise would be to urge on me a principle that very few epistemologists of any stripe have been inclined to accept, viz., that I am only justified in believing *p* if I *know* that I am justified in believing *p*.

Still, I will answer the question, because I think the answer is interesting. There are two ways that I could acquire good reasons to believe that my cognitive machinery is reliable. One way is direct, and the other is indirect, but both are empirical. The first, direct way, is to investigate the character of the cognitive machinery itself. Such an investigation might yield surprising results. Although I can perform inferences *a priori,*
I cannot necessarily discover the character of the cognitive processes that underlie those inferences by introspection alone. That is not to say that introspection cannot yield knowledge, nor that when it does so, the knowledge is not a priori. I am assuming that the “experience” involved in introspection falls on the “inside” side of the etiological divide in terms of which I propose to reconstruct the distinction between a priori and a posteriori belief; hence the deliverances of introspection count as a priori knowledge just when and to the extent that the relevant form of introspection provides reliable access to the processes or thought contents introspected. It is, though, an empirical question for us to answer through scientific investigation whether the particular forms of introspection on which we rely to write philosophy essays about the nature of the human mind and the requirements for knowledge are, in fact, reliable. Surely the data of introspection must be given their due weight, but they can, like other data, be overridden.

Note too, in this connection, that the data of introspection are not unequivocal. Intuitionists do not accept as valid the same class of inferences as classicists, and it is of course a matter of great controversy whether this shows that the connective “not” means something different for intuitionists from what it means for classicists. Vann McGee argued, by displaying counterexamples, that modus ponens itself is not a universally valid rule of inference for the conditional.\(^3\) I cannot here enter the debate about McGee’s examples; I wish only to endorse his methodological conclusion:

The methodological moral to be drawn from this is that, when we formulate general laws of logic [read: when we attempt to characterize the overall structure of our cognitive

machinery], we ought to exercise the same sort of caution we exercise when we make
inductive generalizations in the empirical sciences. We must take care that the instances
we look at in evaluating a proposed generalization are diverse as well as numerous.4

McGee goes on to say that this recognition need not worry those who maintain that logic is “an a
priori science.” It could well be, he says, that we do recognize the validity of an instance of a
valid rule by means of an a priori intuition, without our being able to recognize by means of an a
priori intuition, which exact valid rule it is in accordance with which we are reasoning.

The second, indirect way, of confirming the reliability of our cognitive equipment, and
hence of warranting our belief that our a priori inferences are warranted, is a matter of assessing
the results of the epistemic practices that engage that equipment. Overall epistemic success –
whether it is the everyday success of inferring shape from motion or the world-historical success
of constructing a scientific theory of motion – epistemic success vindicates whatever a priori
elements figure importantly in the epistemic strategies that produce the success. Let me
emphasize once more that this is not to say that warrant for one’s ground level a priori
judgments awaits the outcome of empirical investigation into the reliability of the mechanisms
that produce them. I am warranted in drawing my a priori inferences just in case the mechanism
by which I draw them is a reliable one.5 But what the account does allow for is something that
might bother defenders of the traditional a priori: it allows for the empirical justification of
epistemic practices the cogency of which we nonetheless appreciate a priori. I am not sure that
this should be any more paradoxical than the fact that we can empirically confirm the predictions

4 Ibid., p. 468.

5 I was less clear about this than I should have been in an earlier discussion of this issue
(“Naturalized Epistemology, Morality, and the Real World,” Canadian Journal of Philosophy,
Supplementary Volume 26, edited by Richmond Campbell and Bruce Hunter, pp. 103-137). Thanks to
Richmond Campbell and Joe Levine for raising questions about the account of the a priori given there.
of arithmetic.

In short, I contend that this account allows us to reconstruct, within a naturalized framework, everything that we could reasonably want in the way of a theory of a priori knowledge. *A priori* access is explained by the special etiology of *a priori* beliefs; a *a priori* warrant is conferred on beliefs arrived at *a priori* when and only when the interior processes that give rise to the beliefs are reliable – in accord with the general account of warrant offered by reliabilism. If we want more than that – if we want to be warranted in believing that our *a priori* beliefs are warranted, then empirical investigation is necessary.

It should be clear by now that nothing in a naturalized approach to knowledge prevents one from pursuing the traditional normative questions that seem to many to be constitutive of epistemology. Naturalized epistemology is just as concerned with the question of right epistemic practice as its traditional predecessor. Indeed, the story that the naturalized epistemologist can tell about the *a priori* not only leaves room for norms, but provides the first real clue as to what it might be like to have a rational justification for the acceptance of norms.

II

So the picture is this: we have *a priori* intuitions about such things as logical consequence and cogent empirical reasoning. The more we discover about the reliability of these intuitions, the more reason we have to self-consciously endorse the norms they encode.

But scientific investigation giveth, and scientific investigation taketh away. What happens if and when our empirical investigation into human epistemic strategies fails to vindicate our *a priori* intuitions about what constitutes good epistemic practice? There are two sorts of worries here: the “liberal” worry, and the “conservative” worry. I’ll describe and
address them in turn.

The liberal is afraid that a naturalized approach to normative questions forecloses the possibility of reform: the worry is such an approach makes it impossible to assume a critical stance toward the norms that human beings are discovered to obey, provided these norms work serviceably well. If these are the norms that we in fact endorse, would it not be illegitimately a prioristic to recommend others instead? Doesn’t the naturalized approach mean agreeing that fifty million Frenchmen can’t be wrong?

No. It may be that some defenders of a naturalized epistemology feel duty-bound to canonize current epistemic or scientific practices, whatever they may be, on the bare grounds that this is what we do here, but this attitude is not mandated. Instead, as I see it, naturalized epistemology is concerned with explaining the success of human epistemic practice – but only to the extent that it is successful. There is no reason, therefore, why a serious empirical study of the strategies that we actually employ in everyday reasoning cannot provide evidence – and powerfully compelling evidence – of the need for reform. The now famous work of psychologists like Kahneman, Tversky, Slovic, Nisbett, and Ross shows that in certain predictable situations, human beings who rely entirely on intuition and reflexive judgment reliably fall prey to a variety of “cognitive illusions” – we fail to take account of background probabilities, we assign overly high probabilities to salient possibilities, we systematically neglect certain logical options, and so forth.6 Although such foibles are clearly part of human epistemic life, there is no reason, from a naturalized perspective, to endorse them, and there is every reason to criticize and reform them.

The critical process I envision here is no different from the one described by many

6 For a summary of the relevant work, see Stich (1990) and Stein (1996).
philosophers and historians of science with respect to the development of a variety of scientific
taxonomies, particularly biological ones. (And insofar as this is true, it supports my contention
that there is no sharp division between normative and descriptive study.) Classifications of
animals or natural phenomena often undergo revision in light of empirical discoveries of deep
similarities and differences among animals classified apart or together: whales used to be
thought of as fish (or so the story goes), but a more adequate biology tells us they are not fish,
but mammals. So does a naturalized epistemology tell us that some judgments and inferences
we thought to be sound are in fact fallacious.

But there’s also the second set of concerned citizens: the conservatives. Their concern is
not that naturalized epistemology will forebear from making critical recommendations; quite the
 contrary. The conservatives are worried about a revolution. They fear that a naturalized
approach could end up overthrowing our deepest and most cherished epistemic norms. After all,
these critics may say, on the naturalistic picture, nothing is safe from critique: what counts as
knowledge, what counts as a good epistemic norm – these and all other foundational questions
become matters of the overall goodness of an empirical theory. Our attachment to our norms is
not self-warranting. Rather, the justification of our endorsement of a norm must be instrumental:
a norm is good if, and only to the extent that, its foster our epistemic goals, where these “goals”
are identified, in turn, by a posteriori investigation of our actual cognitive activity, and not by
conceptual analysis or other “first philosophy.”

It’s tempting to dismiss these concerns as alarmist. How real, after all, is the danger?
What traditional epistemologist would want to say that the epistemic norms he or she endorses
do not facilitate our epistemic goals? And what epistemic goals could a naturalized
epistemologist come up with that a traditional epistemologist would eschew? But I think the
conservatives have a point, one that becomes apparent if we look at a different kind of example. The cases mentioned above, in which we discover that human beings routinely engage in fallacious patterns of thinking, are all ones in which traditional epistemic norms are presumed and upheld. But if the naturalized approach is taken fully seriously, it must be admitted that there is a real possibility that traditional norms may instead be called into question.

I have argued, for example, that work in linguistics and cognitive psychology strongly supports the view that our most mundane and fundamental cognitive achievements – acquiring language, recognizing faces, understanding the behavior of our fellow human beings – depend upon our possessing strong native biases of various sorts. Some of these may be simply innate preferences for certain kinds of stimuli – human speech sounds, for example – over others. In other cases, they may amount to theories of particular domains. According one intriguing theory of autism, propounded by psychologists Uta Frith and Simon Baron-Cohen, the disorder is caused by the absence of a “theory of other minds”, an understanding, innate in unaffected humans, of such things as the meanings of various human vocal and facial expressions. (Frith 1989, and Baron-Cohen 1995). These findings cohere with and support the more general point made by Quine and others that human theories in general are vastly underdetermined by sensory evidence, to the extent that we must bring to the task of theorizing an array of tools for paring down the set of hypotheses consistent with our paltry bodies of data, or else flounder forever. All in all, I contend, these considerations show there to be something wrong with a norm of objectivity where this is conceived of as perfect impartiality: it is not simply that human beings are not capable of genuine impartiality (though they probably are not); it is rather that, for creatures of our sort, the implementation of perfect impartiality would be an epistemological disaster.
I have elsewhere argued that this epistemological point – the importance of being partial – has political significance. Acknowledging the ubiquity and necessity of various kinds of pretheoretic biases, I argued, supports the efforts of feminist epistemologists concerned to expose the pernicious effects of an ideology of objectivity. “Dragnet Epistemology”, as I’ve called it, is the often inchoate view of the structure of organized human inquiry, and of science in particular, according to which “objective” research is a straightforward and somewhat mechanical transformation of observed facts into theories, theories that undergo continuous and rigorous empirical testing, and that are rejected if experimental predictions fail to be borne out.

This cartoon view of scientific objectivity carries over, I contend, to other arenas of authorized opinion, including history, politics, and economics. It operates, de facto, to legitimize the opinions of the powerful, and to discredit (often as “biased”) the viewpoints of subalterns. But once the subalterns – feminists, in this case – can show that the features cited as evidence of partisanship in dissident voices are not only present in the mainstream voices taken as paradigms of “objectivity”, but are, in any case, features that are endemic to good epistemic practice, we can show that there are no formal grounds for dismissing the dissonant voices, and, if we are dealing with people of good will, we can refocus critical attention back to its proper focal point – to the substance of the dissident’s claims.


8 This view is often made explicit when science comes under attack, as it is, alas, in my own state of Ohio, where debate is raging about whether to require, allow, encourage or prohibit the teaching of “intelligent design” alongside evolutionary theory. I must say that some of the defenses of science are as embarrassing as the attacks. [Ref to letters to editors] And the rhetoric of Dragnet Objectivity is regularly invoked against the scientific pretensions of the “evolutionists,” who, it must be admitted, hold on to their theories sometimes in the face of recalcitrant data, etc.
It is hardly necessary for you to accept my analysis of political discourse in the U.S., 
circa 2002, because what I want to focus on is the conservative reaction to all this. The 
conservative demands at this point to know how far people like me are prepared to go in 
endorsing bias. Do we envision science as nothing more than the play of partisan forces? Do we 
mean to be saying that everyone should just take things the way they’re inclined to, and make no 
effort to achieve a modicum of disinterestedness? What’s to prevent science from turning into 
mere wishful thinking? Or political advocacy? And by the way, aren’t we the ones who think 
prejudice is a bad thing?

This is the problem that I have elsewhere called “the bias paradox:” the problem that 
progressive complaints about the role of class, gender or other bias in scientific research tend to 
be self-undermining when combined with critiques of the norm of impartiality. What’s really 
needed, both to solve the bias paradox and to answer the conservative critic, is some principled 
way of distinguishing the good biases from the bad biases. When I first wrote about this 
problem, I argued that what naturalized epistemology had gotten us into, naturalized 
epistemology could get us out of – that is, the same naturalized approach that disarmed any 
general attack on bias, could supply a principle for making the needed invidious distinctions. 
The principle I had in mind was this: the good biases are the ones – like the innate tendency to 
develop stereotypes of local flora and fauna – that facilitate the construction of theories that are 
true, whereas the bad biases are the ones – like racial stereotypes – that lead us in the opposite 
direction, or else take us nowhere at all.

Karen Jones, however, has argued that this approach to the bias paradox is unsatisfactory, 
because it leaves the norm of truth too disconnected from the other norms that constitute our 
conception of good epistemic practice (Jones 1998 and personal correspondence). It is not only
the norm of impartiality that seems now to be only contingently connected with truth-tracking; the problem arises for rationality, as well. To this extent, then, Jones’s objection to my reliabilist approach to the bias paradox echoes general complaints that have been lodged against reliabilist accounts of justification *tout court*. As Boghossian puts the point, reliabilism fails “to connect with a thinker’s responsibility for his cognitive practice.” ⁹ To make truth a sort of “master virtue” as reliabilism does suggests that the norm of rationality is *only* instrumentally justified, that it might be legitimately abandoned in any case where some other strategy for forming beliefs leads more reliably to the truth. But the idea that rationality is simply one trick among many in a grab bag of epistemic instruments available to us makes hash of the very notion of epistemic *agency*. Commitment to rationality involves, among other things, a norm that bids us make our reasons transparent to ourselves as we reason – arguably that is what reasoning *is*.¹⁰ To view ourselves as devices that simply undergo different forms of “registration” – with this reasoning thing being one form among others – is, arguably, to give up a conception of knowing as an *activity* at all. Thus the problem Jones is raising can be represented this way: how do we

---


¹⁰ For a probing discussion of the assumption – standard in the epistemology literature -- that justification requires some kind of transparency condition, see Henderson & Horgan (2000a). Henderson & Horgan argue persuasively that justification cannot be made fully transparent because of in-principle limits on computational size in human cognition, and discuss the consequences of this fact for traditional conceptions of epistemology. They also endorse, however, my contention that intuitions requiring transparency must somehow be accommodated if we are to retain the core of our pretheoretic notions of knowledge and justification.
integrate a normative conception of epistemic agency, according to which we ought to deliberate on the basis of considerations we can discern to be evidentially relevant, with a naturalized understanding of ourselves as creatures whose finitude entails that we cannot get by epistemically without shortcuts and tricks of all kinds, many of which would not survive scrutiny by traditional epistemological lights?

II

I want to approach this problem indirectly. Very indirectly. I want to look briefly at ethics. I suggest that the details of the human moral situation raise the same kinds of challenge to the ethical norm of impartiality as the human epistemic situation raise for the epistemic norm of impartiality, and that a naturalized moral epistemology therefore faces an analogue to Jones’ problem. Consideration of the two sets of problems together will, I hope, be doubly illuminating.

Here is the problem in the moral domain, stated baldly: given the facts of our embodied human lives, the constraints of impartialist moral theory seem impossible to satisfy. I do not mean just that our spirits are willing but our flesh is weak – that is, I do not mean to be calling attention to the fact that we often do not do what we think we ought to do. I mean, rather, that the demands of any moral theory that bids us take a disinterested and fully general view of the moral issues that confront us, are demands that creatures who are embodied as we are cannot meet and still be moral agents. The situation, as I see it is pretty precisely parallel to the epistemic case: just as our ability to gain knowledge of our world, given the constraints of our physical finitude, depends upon our not being epistemically indifferent to all the logically available options, so too does our ability to relate morally to each other – to exercise a moral quality of concern toward other people – depends on our not in fact according the same kind and
amount of moral concern to all others at all times.

This theme – ethics on the human plane requires partiality – is one that has been sounded by feminists ethicists, but it has been defended with equal vigor by mainstream ethical theorists. Bernard Williams, for example, argues that a friend who pays a sick call to another friend is not acting properly as a friend if the visit is motivated by a sense of moral duty – whether this is determined by a maxim check against the categorical imperative, or a quick calculation of the effect of such a visit on the well-being of the human population as a whole. (Williams 1982) The proper feelings of friendship, if present, preclude such considerations. A proper friend simply wants to cheer up a suffering friend – it is the suffering of that particular friend that prompts the visit, not the Moral Law or the Principle of Utility, and anyone who appealed to either of the latter to explain her motives would be thought monstrous.

The element of this and other critiques of impartialist moral theories is this: there is moral value in certain kinds of partiality. It is morally admirable for a friend to be motivated simply and directly by the needs of that particular friend, or for a lover to side with a lover in a dispute, or for a mother to have faith in the innocence of a child charged with a crime. We call such attitudes “loyal.” (From an impartialist perspective, it is puzzling how loyalty could be a virtue: if your lover is in the right, then you should side with her for that reason – if she’s in the wrong, then you shouldn’t side with her at all. There seems to be no scope for loyalty, per se.)

Now if we were to take a naturalized approach to moral epistemology – if we were to allow our ethical theories to be informed by a realistic view of the details of our moral practices, then we would have to make some accommodation within our ethical theory for these facts about the salutary effects of certain kinds of ethical partiality. But then our naturalized moral epistemology faces the same conservative challenge as confronted us in the general case,
complete with an ironic twist: how are we supposed to account for the moral value of partiality, without throwing justice out the window? How do I give value to my child’s well-being in an appropriately partial way, while staying mindful of the needs of other children that happen not to be mine? How do I negotiate conflicts between loyalty to friends and commitment to abstract principles of either a moral or an epistemic kind?

If we modify our conception of morally permissible behavior by a recognition of our de facto moral judgments, which apparently condone a great deal that our stated principles appear to condemn, we must ask what sense we are to make of our own moral agency. How are we to represent to ourselves the reasons according to which such “tolerated” behavior is deemed morally permissible or even morally good, if our moral agency is partly constituted by our binding ourselves to the norms that these problematic moral judgments belie?

Once again, the picture that seems to be presented by the naturalized approach has our agency “dissolving” into an agent-less dynamical system that simply responds to a variety of forces: a principle here, a personal attachment there. But this picture is repugnant. Impartial moral principles – for example – are not just “soft constraints” to be optimized alongside lots of others. Nor are we passive moral registers, mere devices for the detection of the good and the right. No – we are moral agents, which mean that we represent our moral situations to ourselves, and deliberate about how to proceed. At least this is how it appears to us, and if it turns out not

---

11 This is not to speak against connectionist theories of the acquisition of moral knowledge (see, for example, Paul Churchland’s proposal in (Churchland 1989). I have other things to say against such theories, but these other things have to do with the shortcomings of connectionist theories in general, and not with problems for applying connectinism to moral learning in particular.
to be true, it will mean, probably, that we have lost moral value itself.

So this is the problem that naturalized moral epistemology must solve, analogous to the one that naturalized epistemology in general must solve: find the norms that save the appearances, where the appearances include our sense of our selves as active knowers and moral actors. Or rather, what I should say, is that any epistemology must solve these problems, and if any can, my bet is that it will be a naturalized one.

With this parallel in mind, let’s look back in some more detail at the empirical debate about human rationality, where the character of the problems looked to be “reformist” in nature, and see what lessons we might derive that might be applied in the more troubling epistemic cases.

Kahneman & Tversky and others have demonstrated that the judgments ordinary people make in certain circumstances do not conform to principles of “best reasoning” – they will completely ignore base rates, for example, and will violate the principle of conjunction (the probability of ‘p & q’ cannot exceed the probability of ‘p’). What can we conclude from such findings? Before indulging in wholesale self-depredation, let’s note two facts: First, notice that endorsement of the violated principles is also forthcoming under certain other circumstances. My evidence here is largely anecdotal, but many subjects will endorse the violated principles when they are presented abstractly – I’ve never seen anyone try to deny the conjunction principle, for example – and many people can be brought to see that they have made mistakes. Second, note that the violated principles are themselves extractable from human epistemic practice – they must be, since it is human beings who have discovered them. I do not mean to be glib about this: the point is that the principles were not just found in a book labeled “follow me if you want to live” – they were abstracted from reflective judgments of probability and validity, or
inferred from principles so abstracted, where the process of inference itself provides further evidence of human commitment to the practice.

If we take the evidence for human beings’ commitment to various principles of rationality and set it beside the evidence that they often violate these principles, the picture that emerges suggests that we are looking at a competence/performance phenomenon – that logic and probability theory are part of human epistemic competence, which, in interaction with other systems like memory and perception, produce sometimes imperfect performance.\textsuperscript{12} The trick then becomes getting an accurate delineation of the environmental circumstances in which the interacting mechanisms produce error. This may be hard to do – sets of circumstances may have features in common that are not obvious, and others that look similar may turn out to be crucially different. The work of psychologist Stephen Ceci and others shows that human “intelligence” is a much more complex phenomenon than either psychometrical or philosophical models presume. (Ceci 1996) People’s skill at deploying many cognitive abilities that might reasonably be assumed to context-independent has been shown to vary markedly from setting to setting. Murtagh (1985), for example, showed that the majority of supermarket shoppers instructed to optimize volume for price, without aid of either posted unit pricing or calculator, were able to make the correct choices. When the same subjects were later given the MIT test of mental arithmetic – a measurement, presumably of the subject’s skill at performing the very mental operations they had just employed – there was found to be no relation between test performance and shopping accuracy.

On this picture, deviations from norms evident in actual human behavior are still analyzed as errors: logic and probability theory are right, and our judgments are wrong, and the

\textsuperscript{12} I acknowledge that this diagnosis is not universally accepted. See ref
only question is, how did we goof? Typically, the answer will advert to some feature of our psychology or our circumstances that limits or interferes with the good judgment embodied in our epistemic competence. The analogue in the moral case would be the situation in which some deeply endorsed value should be implicated, but is not, due to the “interference” of some other faculty. Emotion and affection, of course, have been held by many philosophers (and many ordinary people, by the way – it’s not right to lay this all at the hands of Enlightenment intellectuals) to be the main culprits. But this result will be deeply unsatisfying to the critics of impartialist and rationalistic moral theory: these critics do not think it’s correct to view affective attachments and feelings as amoral, much less as anti-moral. At the very least, they want it to work out that many instances of partiality are morally laudable. And I think that’s right.

The more precise analogue to the really troubling epistemic situations may lie, then, not in the cases where people are properly viewed as committing an error, but in cases where an apparently irrational decision-method turns out to work better than the method suggested by classical principles. Many such cases are described by Gerd Gigerenzer and Daniel Goldstein: they note, for example, the widespread reliability, across a number of domains, of a strategy they call the “recognition heuristic:” when asked which American city is larger, San Diego or San Antonio, 62% of U.S. college students in the study gave the correct answer (San Diego), as compared with 100% of their German counterparts. Gigerenzer’s explanation of this discrepancy is that the German students, who know much less about medium-sized American cities than U.S. students, have likely never even heard of San Antonio, and chose the one city of the pair that they recognize. Since it is, in fact, more likely that a foreign city you recognize will be larger than one that you don’t, reliance on the rule of thumb that bids you go with the one you recognize will serve pretty well. (Goldstein and Gigerenzer 1999, p. 43)
The recognition heuristic, obviously, is not going to be completely reliable -- it depends for its success on a variety of contingent conditions. The users of the heuristic must, for example, know a little, but not too much about the target domain. In this case, the U.S. students cannot use the recognition heuristic because, paradoxically, they know too much about U.S. geography. Because they are familiar with too many U.S. cities, recognition will not serve properly as a filter. On the other hand, if the German students had known nothing about U.S. geography, or if they had been able to recognize only the names of three or four U.S. cities, their recognition filter would have been too crude. Gigerenzer calls such heuristic strategies, which depend for their reliability on contingent but stable features of the environment, “ecologically valid.” Lest you think that such strategems could have only limited value in real-world decision-making, consider this result, reported in Borges, et al. (1999): stock portfolios constructed by polling 360 Chicago pedestrians about which stocks they recognized significantly outperformed portfolios constructed either randomly or by the advice of financial experts.

Here’s another example of an ecologically valid heuristic: In a study of ER diagnosis procedures, it was found that a simple dichotomous choice test (or “classification and regression tree” -- CART) that looked at a maximum of three factors did better at classifying patients as high- or low-risk for heart attack than a classical multiple-regression model that took account of 19 known risk factors. (Gigerenzer 2000, and Gigerenzer and Goldstein, 1999, p. 91) Why is this so? An analysis of the contribution of each of the factors to the explanation of variance reveals that the first factor – minimum systolic blood pressure over a specific time interval – explains a whopping percentage of the variance. After that, the drop-off for the other two factors incorporated into the test is extreme, and after the third factor, all the remaining factors
contribute almost nothing. In situations with this structure, a model that takes account of the neglibile factors is actually more likely to mis-predict in new cases than a model that neglects them. (New motto: Neglect the negligible.) As Gigerenzer points out, the more information a model is required to account for, the less robust it is – the less likely it is to work well in new cases. The problem, intuitively, is that the more information we get about a limited set of cases, the more idiosyncratic our data may become. (Gigerenzer 2000) This point is, I think, identical to the point that philosophers of science have been making, at least since Quine, about the trade-off between simplicity of hypothesis and empirical adequacy. (Quine and Ullian, 1978).

What do these cases show? They might be taken to show that the norms embodied in classical statistical methods are wrong – after all, statistics is meant to be a guide to the best ways to draw conclusions on the basis of limited information. But I think this conclusion would be unwarranted. The cases do show that the utility of the models generated by classical principles are limited by empirical factors, not only by psychological factors that make utilizing the models difficult for human beings – we now have electronic prostheses that obviate those problems – but also by facts about the distribution of information. The more successful strategies in these cases are not, therefore, irrational, and no longer appear to be so once we take the circumstantial facts of our epistemic situation into account. What the existence of ecologically valid strategies shows is not that the norms employed by the classical models are wrong or invalid, but rather that the circumstances of ecology make certain shortcuts better routes to the end goal of predictive validity than the classical computations.

But the availability of an explanation in classical terms of the successfulness of these

---

13 These data, from Breiman et al. (1993), were reported in Gigerenzer (2000). See also Gigerenzer (1999), p. 91.
strategies shows that the norms have not been flouted at all – rather, what’s been shown is that
the best way to conform to certain norms of (in this case) probability estimation is *not to compute directly in terms of the norm*. This is the lesson that heuristics teach us across the board: aiming directly at what you want is not always the best way to get it. Utilitarians have long recognized this in arguing for the distinction between taking the principle of utility as a constitutive norm, and treating it as a decision procedure.

The worry about reliance on instrumentally justified heuristics was that we risked a loss of transparency that would threaten our epistemic agency. These examples show us, though, that even if a certain non-transparency is recommended in *practice*, the transparency is restored at the meta-level, where the norms return to form part of the explanation of the heuristic’s success. This points out two important facts about norms that will apply in the moral case: first, norms can play an explanatory role even when the norms are not explicitly invoked in the procedures whose success they explain. Second, the fact that norms can play such an explanatory role tells us something about the *value* of the norm itself. Let me say a little more about this point.

A naturalized approach counsels us to seek accounts of the value of our norms – unlike a prioristic epistemology, it tells us not to take norms for granted. We need to ask how the adoption of certain norms contributes or would contribute to the goals toward which those norms are oriented. In some cases, it may turn out that the value of a norm is purely instrumental. It might well be argued, for example, that epistemic objectivity, at least considered as a practical norm to be employed in everyday reasoning, has just this character – the removal of partiality or bias is valuable only insofar as it functions to remove idiosyncracy, and idiosyncracy is problematic only insofar as it leads us away from the truth. To the extent that biases represent stable contingencies of the domain under investigation, it will be inefficient, relative to our
immediate epistemic goals, to try to reason without them. More pertinently, nothing of epistemic value is lost if we indulge “biases” of the right sorts – this is, indeed, what justifies our taking a rather cavalier attitude toward departures from objectivity in many everyday cases.

At the same time, a look at the role of norms can also explain why objectivity cannot be so blithely cast aside as a regulative norm. Our broadest and most general epistemic goals involve seeking a kind of epistemic flexibility, an independence from any particular circumstance. We are creatures whose questions and concerns (we presume) far outstrip the epistemic challenges endemic to our ancestral environments, and so we cannot rely as heavily as less ambitious creatures do on being situated in epistemically propitious circumstances. The abandonment of a commitment to objectivity -- construed now as a background imperative to assess the contribution of one’s situation to one’s ability to know -- would mean putting ourselves at epistemic risk in novel circumstances, as relinquishing epistemic projects, like science, not tied to our immediate animal needs.  

In either case, the norm of objectivity is justified instrumentally – it is a good norm, given its contribution to our epistemic goal of seeking useful or meaningful truth. But there is a different kind of justification that a norm can be given. Let’s look back at the norm of impartiality in the moral realm.

Even if we do better in our daily moral practice by indulging at least some feelings of partiality than we would do by adhering strictly to impartialist moral principles, it still strikes me that the norm of impartiality must be accorded a role in our moral life. The reason is that the norm of impartiality expresses something, something extremely valuable and important:

14 See Henderson & Horgan (2000b) for a discussion of this point: the value of, as they put it “practicing safe epistemology.”
commitment to the norm of moral impartiality expresses our commitment to the equal moral value of every human agent. Commitment to this norm appears to me to signal the same kind of extension -- or, if you don’t mind, transcendence -- of our instinctive animal sympathies as science and other self-conscious epistemic practices represent with respect to our evolved epistemic instincts. To treat another being morally is to go beyond sympathetic reaction or instinctive response; it is to appreciate the morally valuable characteristics that the being possesses, and to treat these characteristics as reasons for making a moral response. Moral impartiality is the recognition that these characteristics are present in many, many more beings than the ones to whom we happen to have personal connection; commitment to this norm, then, is commitment to the moral point of view.

Something similar can and should be said about the role of rationality in our epistemic practice. Certain aspects of rational practice – reasoning according to the laws of logic, for example – are certainly instrumentally justifiable, since they facilitate truth-tracking. But rationality in all its guises has another, expressive function as well. Rationality, as I said earlier, is a norm that bids us make our reasons transparent to ourselves as we reason. Since it is the active and self-conscious consideration of reasons that makes one an epistemic agent, the norm of rationality can be said to express our conception of what it is to be an epistemic agent; and to endorse the norm is to express one’s commitment to the value of such agency. The difference this makes to our attitude toward the norm is important. Whereas our commitment to objectivity is conditional – we should follow the norm unless and until it fails to get us the kinds of truths we want – our commitment to rationality is absolute. Well, not quite absolute: to be committed to rationality is to be a certain kind of knower – an active and self-conscious seeker of knowledge. It must be acknowledged that it is not impossible for a human being to relinquish
their commitment to rationality, and hence not impossible for a human being to cease to be an active and self-conscious knower. I thus cannot say that there are no conditions under which a human could cease to value rationality; I can claim only that it is deeply regrettable if it happens. Epistemic agency, I contend, is a profoundly valuable thing: a mind is a terrible thing to waste. So the loss of the norm of rationality would mean, in and of itself, the loss of something of value.

I have developed a parallel with moral norms in order to bring out different ways in which a naturalized approach to normativity can yield surprising, but ultimately satisfying results. But there is more than a rhetorical connection between the two realms. I do think that rationality, and indeed, self-conscious rational reflection, may be causally necessary for human beings to obtain a great deal of the knowledge we manage to obtain. But I do not think, as befits a reliabilist, that the availability of rational grounds is constitutively necessary for warranted belief. The role I reserve for norms of rationality adverts to our values – our valuing of and respect for, our own agency and that of others. This respect is evidenced by our practices of giving and accepting reasons in our epistemic economies. I may have a faculty of intuition that gives me direct access to the truths of mathematics; I still recognize an obligation, if I value rational agency, to offer you a proof rather an oath when you ask me why I believe some particular theorem. Similarly, my religious experience may be the reliable grounds for my belief in God; I still feel bound, if I value rational agency, to respond to the atheist’s objection to my position.

Perhaps this makes epistemic norms into moral norms. That’s OK with me. Then surely no one could say that a naturalized epistemology has no room for the normative.