

“Reflective equilibrium and underdetermination in epistemology”

Jared Bates, Indiana University Southeast

*Acta Analytica* 19/32 (2004): 45-64

Abstract: The basic aim of Alvin Goldman’s approach to epistemology, and the tradition it represents, is naturalistic; that is, epistemological theories in this tradition aim to identify the naturalistic, nonnormative criteria on which justified belief supervenes (Goldman, 1986; Markie, 1997). The basic method of Goldman’s epistemology, and the tradition it represents, is the reflective equilibrium test; that is, epistemological theories in this tradition are tested against our intuitions about cases of justified and unjustified belief (Goldman, 1986; Markie, 1997). I will argue that the prospect of having to reject their standard methodology is one epistemologists have to take very seriously; and I will do this by arguing that some current rival theories of epistemic justification are in fact in reflective equilibrium with our intuitions about cases of justified and unjustified belief. That is, I will argue that intuition underdetermines theory choice in epistemology, in much the way that observation underdetermines theory choices in empirical sciences. If reflective equilibrium leads to the underdetermination problem I say it leads to, then it cannot satisfy the aims of contemporary epistemology, and so cannot serve as its standard methodology.

The basic aim of Alvin Goldman’s approach to epistemology, and the tradition it represents, is naturalistic; that is, epistemological theories in this tradition aim to identify the naturalistic, nonnormative criteria on which justified belief supervenes (Goldman, 1986 22-4; Markie, 1997 38). The basic method of Goldman’s epistemology, and the tradition it represents, is the reflective equilibrium test; that is, epistemological theories in this tradition are tested against our intuitions about cases of justified and unjustified belief (Goldman, 1986 66; Markie, 1997 38-9). I will argue that there is an acute problem with the relation between contemporary epistemology’s basic aims and its basic methods. I will argue here that its basic methods do not serve its basic aims.

Suppose for now, as I will later show, that the standard methodology of recent epistemology is indeed the reflective equilibrium test. Now, what if multiple, mutually incompatible theories of justification are in reflective equilibrium with our epistemological intuitions? Rather than accept the incompatible theories, we should reject the standard methodology. I will argue that the prospect of having to reject their standard methodology is one epistemologists have to take very seriously; and I will do this by arguing that some current rival theories of epistemic justification are in fact in reflective equilibrium with our intuitions about cases of justified and unjustified belief. That is, I will argue that intuition underdetermines theory choice in

epistemology, in much the way that observation underdetermines theory choices in empirical sciences. If reflective equilibrium leads to the underdetermination problem I say it leads to, then it cannot satisfy the aims of contemporary epistemology, and so cannot serve as its standard methodology.

## 1. THE BASIC AIMS AND METHODS OF EPISTEMOLOGY.

It has to be granted that the aim of much work in contemporary epistemology is to give epistemic norms like justification naturalistic accounts. More specifically, theories of epistemic justification aim to tell us what nonnormative, or purely descriptive, features justified belief supervenes on. To do so, a theory must specify, in nonnormative terms, criteria that all and only justified beliefs satisfy. This, I take it is the standard expression of naturalism about epistemic norms (Goldman, 1986 22-4; Kim, 1988 382; Markie, 1997 38).

Even so, there are epistemologists who do not carry their naturalistic aims quite this far, and yet they maintain some legitimate claim to offering naturalistic accounts. Lehrer (1990), for one, imposes no restriction to nonnormative terms in stating criteria of justification, though he does require that the conditions expressed in an analysis be “enlightening” (5-9). He takes it that this goal of the theory of justification is continuous with the goals of natural science, and is moreover quite congenial with even the naturalistic views expressed by Quine (1953, 1969). Presumably, then, an account of justification could be enlightening, in Lehrer’s sense, without immediately satisfying what I take to be the standard view of naturalism in epistemology, if, for example, it contained some residual normative terms. In fact, Lehrer’s own coherentist account, one we will examine later, is just such an account. In such cases, a general naturalistic thesis would hold that the residual normative properties to which such an account adverts would themselves permit of properly naturalistic analyses.<sup>1</sup> So, an account of justification that satisfies Lehrer’s aims can be seen as furthering the standard naturalistic aims, insofar as the number of normative properties to be analyzed becomes simpler. Thus, while the standard view of naturalism in epistemology is quite demanding in virtue of its restriction to nonnormative criteria, we should recognize that some less restrictive aims nevertheless further the general shift toward metaphysical naturalism in epistemology.

Goldman’s project itself makes the naturalization of epistemic justification a two-step process, the first of which relates justification to another squarely normative concept. Since Goldman views matters of epistemic justification as

---

<sup>1</sup> Lehrer himself appears to stand neutral on whether or not the remaining normative terms in his account of justification can be further analyzed into nonnormative terms (1990 127).

matters of permission or prohibition, and since these matters lend themselves well to treatment in terms of accordance with or violation of rules, he adopts the following rule framework for the theory of justification:

S's believing p at time t is justified if and only if S's believing p at t is permitted by a right system of justificational rules (J-rules). (1986 59)

The task of epistemology, then, is to identify the nonnormative or purely descriptive features on which justificational-rule rightness supervenes, the naturalistic properties possessed by all and only correct justificational-rules. Given this aim, it is crucial to see how Goldman proposes to reach it, that is, what methods should be used to decide on a rightness criterion for justificational-rule systems:

The strategy I endorse is best expressed by the Goodman-Rawls conception of "considered judgments in reflective equilibrium." We examine what rule systems would likely be generated by each candidate criterion. We reflect on implications of these rule systems for particular judgments of justifiedness and unjustifiedness. We then see whether these judgments accord with our pretheoretic intuitions. A criterion is supported to the extent that implied judgments accord with such intuitions, and weakened to the extent that they do not. But our initial intuitions are not final. They can be pruned and adjusted by reflection on candidate rule systems. There are other tests of a criterion's adequacy as well. Does it generate any rule systems at all? Does it generate a complete rule system, that is, one that would imply justifiedness or unjustifiedness for all cases of belief and all doxastic attitudes? (1986 66, first emphasis added)

Thus, Peter Markie has attributed the following methodological principle to Goldman-style epistemology, a principle he takes to capture the standard practice of epistemologists:

To see if a criterion of rightness for justificational rules specifies the nonepistemic, factual conditions on which epistemic justification supervenes, we should, among other things, examine the extent to which the justificational rules that satisfy the criterion actually permit, and so endorse as epistemically justified, the most uncontroversial instances of justified beliefs, those we intuitively regard as epistemically justified. (1997 39)

An epistemological theory is to be tested against our intuitions, our classificatory judgments about particular cases of justified and unjustified belief. A theory passes the test by saving our intuitions, that is, by classifying particular cases of justified and unjustified belief according to our intuitive classifications.

As Goldman mentions, not all of our pretheoretic intuitions are authoritative in this way; and here is where his reference to Goodman (1965) and Rawls (1971) is instructive.<sup>2</sup> According to Rawls, before they can be admitted into the base of intuitions with which an equilibrium is to be struck, our initial intuitions must first be filtered: Only those that survive careful and impartial examination should be admitted into the equilibrium base (1971 47-8). Thus, in evaluating principles of justice, it is doubtful that sexist or racist intuitions would enter the equilibrium base since such intuitions are not likely to survive impartial examination, no matter what their initial appeal. And according to Goodman, on presumably rare occasions, we may even decide to give up some of our clearest intuitions for the sake of such things as theoretical utility. His example is our intuitive classification of a whale as a fish; our biological taxonomy is much more powerful and considerably neater if we acknowledge that this is one intuitive classification we are better off without (1965 66). In short, if an intuition is one we are able to negotiate on, then it does not figure into the set of judgments we expect a theory must get right. These I will call our strongly nonnegotiable intuitions.

Implicit in all of this discussion is a further requirement, namely, that the base of intuitions in the reflective equilibrium test be very widely shared. Consider Rawls' remarks about checking the adequacy of principles of justice chosen from the original position:

We can note whether applying these principles would lead us to make the same judgments about the basic structure of society which we now make intuitively and in which we have the greatest confidence; or whether, in cases where our present judgments are in doubt and given with hesitation, these principles offer a resolution which we can affirm on reflection. There are questions which we feel sure must be answered in a certain way. For example, we are confident that religious intolerance and racial discrimination are unjust. We think that we have examined these things with care and have reached what we believe is an impartial judgment not likely to be distorted by an excessive attention to our own interests. These convictions are provisional fixed points which we presume any conception of justice must fit. But we have much less assurance as to what is the

---

<sup>2</sup> The method under discussion is most notably associated with Rawls, who gave it its name; but even for Rawls the method clearly inherits its prestige from Goodman's use of it to justify induction. See Rawls (1971 20n).

correct distribution of wealth and authority. Here we may be looking for a way to remove our doubts. We can check an interpretation of the initial situation, then, by the capacity of its principles to accommodate our firmest convictions and provide guidance where guidance is needed. (1971 19-20)

Not only does Rawls' repeated use of 'we' and 'our' imply that the intuitions that matter must be pretty widely shared, his examples illustrate the same point: It isn't that he has a clear intuition that institutionalized racism is unjust, it's that almost all reflective persons do. By contrast, surely some are very confident in their intuitions about the proper distribution of wealth and authority; but since such intuitions are in dispute, we do not hold a theory of justice directly accountable to them. These I will call our widely shared intuitions.<sup>3</sup>

What does all this tell us about the intuitions that figure in the reflective equilibrium test? For starters, they must be very widely shared among reflective persons. This not only rules out 'minority views,' or idiosyncratic intuitions, it also rules out intuitions about which reasonable people divide among, say, two or three main groups. Moreover, not only must these intuitions be very widely shared, they cannot be easily overturned. In general, the degree to which a theory must accommodate one of our widely shared intuitions is in direct proportion to our general unwillingness to relinquish it.

Goldman's methodology, then, can be more fully characterized as follows: A theory of epistemic justification is to be evaluated against our widely shared and strongly nonnegotiable epistemological intuitions. In the problem at hand, these are our intuitive classifications of particular cases of justified and unjustified belief, classifications that survive careful reflection, are commonly agreed on, and will not likely be given up. A criterion is warranted just in case it is in reflective equilibrium with these intuitions; that is, it is warranted just in case it issues the same (or

---

<sup>3</sup> Goodman similarly intimates that principles are justified by their accordance with shared intuitions. With respect to principles of deductive inference, he states: "A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend" (1965 64, emphasis Goodman's). And with respect to induction: "Predictions are justified if they conform to valid canons of induction; and the canons are valid if they accurately codify accepted inductive practice" (1965 64). The point here is not that principles of inference are justified to the degree that they sanction all and only the inferential practices that Goodman or you or I would accept, but that they are justified to the degree that they sanction the inferential practices that we all, or sufficiently many of us, accept upon careful reflection.

sufficiently close to the same) assessments of justified and unjustified belief as do our widely shared and strongly nonnegotiable epistemological intuitions.

## **2. THE PROBLEM OF UNDERDETERMINATION.**

I will argue that the reflective equilibrium test cannot determine a choice between competing accounts of epistemic justification. This is so because our widely shared and strongly nonnegotiable epistemological intuitions are actually handled quite well by the biggest contenders in the theory of epistemic justification. While these theories certainly differ in their consequences about many cases of belief, they tend to be pretty successful at getting right the intuitions that legitimately figure in the reflective equilibrium test. I think that upon examination it will turn out that these intuitions are pretty meager and that, even taken as a group, they are not robust enough to decide between the going theories of epistemic justification. It is ultimately this underdetermination problem that makes the method of reflective equilibrium ill-suited to the goals of epistemology.

My arguments here are necessarily indirect. One reason is that the number of live theories of epistemic justification, even broadly grouped, is too great to devote attention to each here. One would have to consider, among others, varieties of foundationalism, coherentism, proper functionalism, and reliabilism. My point will stand, though, if I can show that a version of coherentism and a version of reliabilism will both pass the reflective equilibrium test. If that can be shown, then reflective equilibrium will have been shown to be ill-suited to the goals of epistemology, since each of these theories tells us that something quite different is essential to justified belief.

To complicate matters further, the set of our widely shared and strongly nonnegotiable epistemological intuitions is potentially infinite. Our capacity to describe cases of belief that we can assess with confidence is surely open-ended, as evidenced by a multiplication of Gettier-style cases in the theory of knowledge. So, it is practically impossible to prove conclusively that any theory passes the reflective equilibrium test. (This, of course, is a difficulty every positive reflective equilibrium argument faces, not just mine.) And the reasonable response is that, while we cannot conduct the survey required to establish with certainty that a theory passes the reflective equilibrium test, we can nevertheless survey a manageably small class of intuitions in the equilibrium base. And we may reasonably expect that an appropriate sample will map out much of the terrain of intuition, if not reveal all its contours and borders. But here I will not concern myself too much with putting together a sampling of our intuitions about justification that reliabilism and coherentism get right; I will rather try to motivate some general remarks about what sort of intuitions actually figure in an appropriate application of the reflective equilibrium test. I hope to make it clear how vapid those intuitions are, and how

likely it is that an enduring epistemological theory can accommodate them.

Moreover, I will take a look at some particular moves in epistemological debates and show why the reflective equilibrium test does no real work to choose between going rivals in the theory of justification, despite appearances. Before I return to these points, however, let me briefly discuss the coherentist and reliabilist accounts.

Coherentism, according to Lehrer (1990), tells us that an agent's belief is justified just in case it can take all comers on the basis of everything else the agent accepts.<sup>4</sup> On Lehrer's view, a proposition, *p*, competes against a belief, *b*, if and only if *b* is less likely given *p* than otherwise (1990 117-8). Consider my belief, formed in the usual way, that I see a lamp in the corner of the room. One thing that competes against this belief is that there is nothing but a stack of old magazines in the corner of the room. Now, granting some assumptions about other things I already accept, my belief about the lamp is antecedently more reasonable than the claim about the old magazines. I know what lamps look like, after all, and I know that my current experience is as of a lamp, and not just any lamp, but the very lamp I placed there last July. Lehrer would say that the magazine-competitor is beaten on my acceptance system (1990 118). Another competitor is that people sometimes dream that they see lamps. I in fact believe this, so this is one competitor already included in my acceptance system, so it cannot be beaten. But I am confident that I can distinguish between waking experiences and dreams, and I am also confident that I am not now dreaming. This latter statement, conjoined with the dream-competitor, does not compete against my lamp-belief, i.e., does not make it less likely to be true. Moreover, it is at least as reasonable for me to accept both my confidence at distinguishing wakefulness from dreams together with the dream-competitor as it is to accept the dream-competitor alone. The dream-competitor presents a skeptical hypothesis, and Lehrer would say that while it cannot be beaten on my acceptance system, it can be neutralized (1990 125). If all competitors can be so fended off, beaten or neutralized, then my lamp-belief is justified on Lehrer's version of coherentism.

Reliabilism, according to Goldman (1986), tells us that an agent's belief is justified just in case it is the result of a reliable belief-forming process. In other words, justified beliefs are justified in virtue of having been produced by cognitive processes that are good at producing true beliefs.<sup>5</sup> If I look back across the room to see that lamp in the corner and judge that there is a lamp there, then, under some charitable assumptions about my psychological disposition at the time, I am justified

---

<sup>4</sup> This account of coherentism is derived from Lehrer's extended and dialectical development of his theory (1990 115-26).

<sup>5</sup> This account of reliabilism is derived, with some smoothing out, from Goldman's rightness criterion on justification (1986 63) and his reliability criterion on rightness (1986 106).

in my belief. The reason is that visual experience of well-lit, nearby, medium-size objects is a pretty reliable belief-forming process. Others are not so reliable. Wishful thinking, for instance, may be the source of a very optimistic outlook on one's economic well-being over the next year; but wishful thinking, contrary to the hopes of many, does not produce a high ratio of true beliefs. So for all the comfort it provides, the belief in question is not reliably formed, and so is not justified.<sup>6</sup>

It is clear that these versions of coherentism and reliabilism provide distinct accounts of epistemic justification. Reliabilism nowhere references coherence as an essential feature of justified belief, and coherentism nowhere so references reliability. Yet I want to maintain that both these theories handle very well our firmest and least controversial intuitions about particular cases of justified and unjustified belief. I must admit that I am initially tempted to make this point by defying anyone to produce a case of belief about whose epistemic status our intuitions are unambiguous, uncontroversial, and unassailable, but which one of these theories demonstrably gets wrong. I suppose, however, that many think that this has already been done, even many times over. A little later on I will look at some of these purported counterexamples. In the meantime, it will be more helpful to say something positive about the intuitional adequacy of these two theories.

The first thing to say is that both coherentism and reliabilism have undergone sustained criticism and careful refinement over the course of many years of rigorous philosophical engagement. Each has also benefited from both insights and errors of talented predecessors. Lehrer and Goldman have spent their careers fine-tuning their respective theories to take better and better account of our epistemological intuitions. Given this, there is an element of 'no surprise' in the underdetermination thesis I advance. That is, it is no surprise that these theories handle so well our common intuitive classifications of justified and unjustified belief. On the contrary, what would be a surprise is that either of them were to methodically mishandle a systematic portion of our nonnegotiable intuitions.

Consider again the lamp-belief discussed above. I am sitting in a room. Both the room and my cognitive apparatus are in very ordinary conditions. On the basis of a visual experience of the usual sort, I form the belief that there is a lamp in the corner. Here is as clear an example of a justified belief as I can imagine.<sup>7</sup> And, as

---

<sup>6</sup> Not unrelated to this point, I suppose, is the phenomenon of 'depressive realism,' in which dysphoric individuals tend to make more accurate predictions (in certain domains) than do their nondepressed counterparts. For a review of some of the relevant research, see Ackermann and DeRubeis (1991).

<sup>7</sup> Of course, our intuitions about how cases are to be characterized depend on how their details are to be filled out. Our intuition about a case under one description may be withdrawn or reversed when new details are given. And carried to an extreme, no amount of elaboration of a case may suffice for our registering an



indicated in the discussion above, both coherentism and reliabilism give an account of its justification; both, in other words, capture our intuition about the case. Reliabilism tells us that the belief is justified because a reliable belief-forming process, in this case, visual perception, produced it. Coherentism tells us that the belief is justified because, on the basis of the subject's acceptance system, the belief can take out every competitor from flat-out denials to full-blown skepticism.

The lamp-belief in question is pretty obviously justified. And its obviousness is what wins it entry into the equilibrium base in the reflective equilibrium test. Again, the equilibrium base should contain all and only those intuitions that are very widely shared and strongly nonnegotiable (or a representative sample thereof). Thus, other intuitions in the equilibrium base should be as obvious to us as our intuition about the lamp-belief. But such obvious intuitions surely provide scant evidence at best. They include intuitions about obviously justified beliefs like an ordinary person's belief in some elementary statements of arithmetic or recollections about what was for breakfast this morning. They also include intuitions about obviously unjustified beliefs like hasty generalizations and paranoid delusions. Laurence BonJour (2002) begins a recent introduction to epistemology by outlining eleven general sorts of facts we intuitively claim to know, giving specific examples of each that he in particular intuitively knows.<sup>8</sup> Examples include "that I feel an itch in my left thigh ... that I have two hands ... that there are several million people in New York City ... that I had Grape-Nuts for breakfast this morning ... that there was a worldwide Depression in the 1930s ... that my dogs are excited by the prospect of a walk ... that some people are afraid of water ... that I will eventually die ... that  $2+5=7$ " and others as unremarkable as these (2002 2-4). If these are the sorts of data that figure in the reflective equilibrium test, then neither reliabilists nor coherentists have to worry about buckling under to such intuitions. Both reliabilism and coherentism, not to mention many of their sundry challengers, get these lackluster cases in the reflective equilibrium test pretty much right.

Think for a moment what it would mean for an intuition to be less obvious than our intuition about the lamp-belief or those on BonJour's list. It would mean that we feel some room to go either way on it or that in fact reasonable people disagree about which way to go. In either case, though, the intuition will not figure in an appropriate application of the reflective equilibrium test. This is just to say once again that a theory cannot be thrown out for how it handles cases on which we do

---

intuition with absolute confidence. In practice, though, we set such extreme caution aside. Understood in one very natural way, the belief in question above is obviously justified.

<sup>8</sup> BonJour lists these as cases of knowledge. Given the tradition, I take it that these same cases would count as justified beliefs.

not have widely shared and strongly nonnegotiable intuitions, at least, not on the reflective equilibrium test.<sup>9</sup>

In light of these remarks, it is interesting to look at the structure of actual dialectics in the epistemological literature. After all, it appears that epistemologists are using the method of reflective equilibrium both to refine their own theories by bringing them into equilibrium with ordinary intuitions and to reject rival theories by illustrating their counterintuitive consequences. I think that closer examination will make clear that something else is really going on: Epistemologists tend to appeal to uncontroversial intuitions about very ordinary cases when refining their own theories and count this as somehow justifying a choice of their own theories over others; on the other hand, they tend to appeal to weak or controversial or idiosyncratic intuitions about extraordinary cases when they launch objections against rival theories. Often, they are able to leave the impression that they are operating within the prescribed standards of reflective equilibrium when in fact they are not. Let me give a couple of examples of each sort of case. Doing so will validate many of the comments made up to now.

To begin, Goldman considers the following case in developing a refinement of his reliabilist theory:

...Humperdink has attended a series of talks on mathematics by a certain Elmer Fraud. These talks are not under the auspices of any certified educational institution, and Humperdink has been warned that Fraud has no credentials in mathematics. Humperdink hears Fraud enunciate numerous principles and algorithms, almost all of them defective. Nonetheless, being a complete novice—and a gullible one at that—Humperdink blindly accepts and applies them all. In one case, however, Fraud happens to teach a perfectly correct algorithm. Humperdink internalizes this one along with the others, and applies it to a relevant class of problems. In using this algorithm to solve a problem, Humperdink gets the right answer and forms a true belief in the answer. This belief is the result of a reliable process, namely, the algorithm ... Clearly, though, Humperdink should not be credited with knowledge. (1986 51-2)<sup>10</sup>

---

<sup>9</sup> Of course, one might think that controversial intuitions can appropriately figure in the reflective equilibrium test. But then the method of reflective equilibrium has the unseemly consequence that you will only convince others who already agree with you. That is, a theory's passing the reflective equilibrium test should provide no reason at all for those who disagree with the controversial intuitions.

<sup>10</sup> Goldman considers a parallel case to make essentially the same point later on (1986 91).

The case allows Goldman to clarify the notion of reliability in a crucial way. Namely, there may be a reliable method for arriving at beliefs in a certain domain, but if the method itself is not selected in a reliable way (as in Humperdink's case), then beliefs formed through its use will not be justified. So there is first- and second-order reliability. Where beliefs are formed using acquired methods, reliability at both levels is necessary for justification. With the distinction between first- and second-order reliability in hand, reliabilism can account for the intuition that Humperdink's belief in this case is unjustified: The algorithm Humperdink adopted was a reliable method for reaching the relevant sorts of belief, but the way Humperdink adopted it was not reliable. So the beliefs he formed were unjustified on reliabilism.

Lehrer, too, appeals to obvious cases in refining his theory of justification. He imagines the following scenario: Someone is looking at a zebra in a zoo and forms the belief that he sees a zebra. We may suppose that Lehrer's zebra-belief is as unspectacular as our earlier lamp-belief. And so it is as obviously a justified belief. Nevertheless, at this point in its evolution, Lehrer's coherentist theory provides no means of coping with skeptical competitors like the fact that people sometimes dream that they see zebras; that is, up to here, his account requires that all competition is beaten on an individual's acceptance system. Consideration of this case leads Lehrer to modify a tentative statement of his account. A subject's belief is justified if all its competitors can be either beaten or neutralized. On the revised account, our intuition that the zebra-belief is justified is then easily captured (1990 119-26).

Here we have seen two epistemologists consult their intuitions about cases to strike an equilibrium and thereby promote their theories. But the intuitions they cite have two very important features in common: First, they are among our strongly nonnegotiable, widely shared intuitions; and second, every other epistemological theory being defended in the literature today can accommodate them. That each captures our intuitions about these respective cases is of no advantage to it over its rivals.

Appeal to intuition in rejecting a theory is at least as common. Consider Lehrer's tale of Mr. Truetemp. Truetemp, without knowing so, has had a device implanted in his brain that at regular intervals produces correct beliefs in him about his brain's temperature. Thus, every hour on the hour, Truetemp forms the belief that his brain's temperature is 98 degrees (just in case his brain's temperature really is 98 degrees). What is interesting about Truetemp is that his "doxatemp" beliefs are formed by a reliable belief-forming process, and so ostensibly satisfy the reliabilist's criterion of epistemic justification. But Lehrer says that Truetemp's doxatemp beliefs are not epistemically justified. Truetemp, after all, has no idea why he believes that the temperature of his brain is 98 degrees—he just believes it. The case seems to show that reliability is not sufficient for justification. Lehrer seems to have produced

a case of belief, Truetemp's doxatemp belief, that reliabilism classifies as justified and that ordinary intuition classifies as unjustified.<sup>11</sup>

In another example of appealing to intuitions in order to reject a theory of justification, Plantinga (1993a) devises the following case against reliabilism, the case of the Epistemically Serendipitous Lesion. Plantinga imagines himself the unlucky victim of a rare form of brain lesion:

As a result of this lesion (and we can stipulate, if we like, that it develops prenatally), most of my beliefs are absurdly false. It also causes me to believe, however, that I am suffering from a brain lesion. This belief of mine, pathologically caused as it is, is clearly one that has little or no warrant. But there will be a cognitive process, in Goldman's sense, whose output is this belief; and we may suppose that this process—call it 'P<sub>1</sub>'—occurs only in conjunction with a lesion of the sort in question. So the result of adding

R<sub>4</sub> P<sub>1</sub> is permitted

to a right system of [justification]-rules will itself be a right system. Hence Goldman's account yields the conclusion that this belief is justified and, indeed, that under these conditions I *know* that I am suffering from a lesion. (1993a 201)

The case of the epistemically serendipitous lesion is supposed to provide an intuitively unjustified belief that comes out justified on reliabilism. Since reliabilism gets this intuition wrong, Plantinga concludes that there is something wrong with reliabilism.

The last two cases, Lehrer's case of Mr. Truetemp and Plantinga's case of the Epistemically Serendipitous Lesion, deviate from the earlier two cases in some interesting ways. To see how, recall what gains an intuition entry into the equilibrium base in the reflective equilibrium test. To get into the equilibrium base, an intuition must be uncontroversial, that is, very widely shared among those who have considered the case carefully. It must also be strongly nonnegotiable, that is, so firmly held that, other things equal, we would rather reject contrary theories than give up the intuition. Do our intuitions about Mr. Truetemp and the lesion-victim qualify as uncontroversial and strongly nonnegotiable intuitions by these standards? They qualify as neither.

---

<sup>11</sup> Lehrer initially put forth the tale of Mr. Truetemp as a counterexample to externalism in general and to reliabilism in particular (1990 165-6); he later resurrected Truetemp as a counterexample to Plantinga's proper functionalism (1996 31-3).

One salient feature of the two cases is that they present only very remote and unfamiliar possibilities. None of us will ever meet a Mr. Truetemp or succumb to a lesion of the sort Plantinga describes. The strangeness of the cases alone, it is reasonable to think, should attenuate any convictions we may have about them. This is not at all to say that a theory of justification is not accountable for hypothetical, or 'merely possible,' cases. A theory is a theory, after all, and as such must handle all the possibilities correctly. The question is, how, without use of the right theory, we know the correct assessment of a given possibility. The epistemic status of any hypothetical belief is surely not transparent. Our access to its epistemic status is not like our access to our own inner states or to the colors and shapes and sounds of things around us. This suggests that, when the time comes around to negotiate down on our intuitions (and, that time will come around), then intuitions about only distant and bizarre possibilities (insofar as we have any) surely belong in the first round of cuts.

But even to say this much about the cases seems to grant that we have firm and abiding intuitions about them in the first place. I doubt this is so in either case. Lehrer's intuition about Mr. Truetemp's doxatemp beliefs, while not entirely idiosyncratic, is not entirely uncontroversial either. Some find it perfectly intuitive that the doxatemp beliefs are justified; after all, Truetemp forms the belief that his temperature is 98 degrees if, only if, and because his brain really is 98 degrees. Nor is Plantinga's intuition clear and obvious. By hypothesis, the subject in Plantinga's case is supposed to have a congenital brain lesion that leaves him with mostly absurdly false beliefs (or better, as we shall see, belief-like states). But on a widely accepted view of intentionality, one necessary condition for being an intentional system is to track mostly truths.<sup>12</sup> The lesion-victim doesn't do this. On this sort of view, then, our unlucky lesion-victim falls outside the scope of epistemic appraisal because he falls outside the scope of intentional description, and so Plantinga's purported counterexample to reliabilism fails.

I think this all points to a problem with the standard practice in epistemology, the practice of appealing to intuitions in reflective equilibrium. Reflective equilibrium cannot justify one epistemological theory over its rivals because our shared and strongly nonnegotiable intuitions underdetermine a choice between competing theories. Since showing that a favored theory handles very well our least controversial and least negotiable epistemological intuitions shows nothing that rival theorists cannot also show, epistemologists often seek to depose rival theories by drawing out consequences that they then claim to find counterintuitive. I think that these are often acts of desperation, and that they have to some extent hidden the underdetermination problem from view. But once we see that the intuitions these sorts of argument tap are not allowed into the reflective equilibrium test (as in the

---

<sup>12</sup> See Dennett (1978 17-8), for just one example of such a view.

cases cited above), not only do the arguments fail, the problem of underdetermination becomes clear.

### 3. IS REFLECTIVE EQUILIBRIUM THE STANDARD PRACTICE?

The underdetermination problem I outline is a problem only if reflective equilibrium is the standard methodology of contemporary epistemology, and only if it cannot be easily skirted around. So, is Goldman's proposed strategy of using reflective equilibrium really representative of a more general epistemological tradition? Appealing to intuitions about cases either to refine or promote a favored theory, or to criticize or reject an opposing theory, is hardly more common in other areas of philosophical inquiry than it is in epistemology. We have already encountered some concrete examples of this standard practice. But it may well be thought that the reflective equilibrium test is not all there is to the standard practice. On the principle Markie attributes to Goldman, quoted in section 1, "we should, among other things, examine the extent to which the justificational rules that satisfy [a proposed] criterion actually permit, and so endorse as epistemically justified, the most uncontroversial instances of justified beliefs, those we intuitively regard as epistemically justified" (1997 39, emphasis added). What "other things" should we do? Markie doesn't say, here, but he seems to be taking his lead from another passage quoted earlier from Goldman: "There are other tests [i.e., tests other than reflective equilibrium] of a criterion's adequacy as well. Does it generate any rule systems at all? Does it generate a complete rule system, that is, one that would imply justifiedness or unjustifiedness for all cases of belief and all doxastic attitudes?" (1986 66, emphasis Goldman's).

Goldman and Markie appear to think that epistemologists should and do bring to bear considerations other than reflective equilibrium on their selection of criteria of justification. Yet the only further tests mentioned by Goldman have us do the following: (1) Check whether a proposed criterion generates any rule system at all; and (2) if it does, then check whether the rule system is "complete." However, these additional tests can be shown to be impotent vis-à-vis the reflective equilibrium test. First of all, if a criterion doesn't generate any rule system at all, then, trivially, it cannot pass the reflective equilibrium test articulated in Markie's principle. That is, if no rule system satisfies a proposed criterion, then it cannot issue any judgments about the epistemic status of beliefs, and so its judgments cannot be evaluated by intuition. So this test, in conjunction with the reflective equilibrium test, is superfluous. The other alternative test Goldman advances is to see whether the rule system generated by a criterion provides a 'total function,' a mapping of some or other epistemic status onto every possible belief. But this test is again trivially satisfied if the criterion is expressed as both necessary and sufficient for justification, as criteria generally are. To see this, recall that Goldman holds that a belief is

justified if, and only if, it satisfies a right system of justificational-rules, and that a system of justificational-rules is right if, and only if, it satisfies a yet-to-be-specified factual condition, C. Simply in virtue of the logical form of these statements, a criterion, C, if it yields any justificational rules at all, will yield assessments of every possible belief. Either a belief satisfies (rules that satisfy) C, or it fails (rules that satisfy) C. If it satisfies C, it has positive epistemic status; if not, not. Of course, it is presumably up to the reflective equilibrium test to determine whether or not C issues the appropriate assessments of beliefs; but whether or not it issues a complete set of assessments is a matter of logical form. So again, this “other” test of a criterion’s adequacy is powerless. Thus, it is difficult to see how these additional tests of criteriological adequacy augment the reflective equilibrium test in any way. A criterion either does or doesn’t issue a set of justificational-rules; if it doesn’t, it will fail the reflective equilibrium test, and if it does, it trivially satisfies Goldman’s urge for a complete function.

It may be suggested that epistemologists can always appeal to the relative parsimony, or other explanatory virtues of competing epistemological theories, in making their theory choices. This suggestion seems to take its lead from a solution to a parallel problem in the philosophy of science. Since observation alone never decides on one empirical hypothesis to the exclusion of all others, scientists base their theory choices on how well empirically equivalent hypotheses stack up on certain explanatory criteria, which typically include parsimony, explanatory scope, conservatism, and fruitfulness. Thus, in choosing between two empirically equivalent hypotheses, scientists may rationally prefer one to the other because it has greater explanatory scope, or it is more parsimonious, or it coheres better with other well established background theories, or it makes more novel predictions, or all of these. Similarly, the suggestion might go, when epistemologists must decide between intuitively equivalent epistemological theories, they can make their choices using these same sorts of explanatory considerations. Since the reflective equilibrium test appears to inherit so much prestige from the model of hypothetico-deductivism in empirical science,<sup>13</sup> it is perhaps only natural to think that the underdetermination problem it faces in epistemology can be solved by appealing to the same criteria used to solve the underdetermination problem in empirical science.

However, although epistemologists can appeal to the explanatory virtues of competing epistemological theories in making their theory choices, this fact doesn’t imply that they actually do make such appeals. That is, it does not follow that it is the standard practice to do so. This is an important point because if explanatory criteria are not standardly brought to bear on theory choices in epistemology, then

---

<sup>13</sup> Some who have held that intuitions in reflective equilibrium play the same evidential role as observation in scientific reasoning include Daniels (1979), Goldman & Pust (1998), Graham & Horgan (1998), and Jackson (1998).

the suggestion that they can be so employed concedes the only argument I am trying to make, namely, that the methods epistemologists actually use underdetermine their selection of one account of epistemic justification over others. I do not think epistemologists in practice appeal to the relative explanatory virtues of competing theories in making theory choices. I have never seen one epistemological account preferred to another because of its relative simplicity or greater explanatory scope. Such appeals are not standard, I suppose because epistemologists do not think they are necessary. Appeal to explanatory criteria is necessary only if competing theories save the phenomena equally well. What recent debates show, as indicated above, is that epistemologists do not think that competing epistemological theories save the phenomena equally well; they do not think their competitors' theories pass the reflective equilibrium test. So since it is not part of the standard practice in epistemology to bring in considerations of explanatory virtue in making theory choices, my underdetermination thesis remains secure.

#### **4. CONCLUSION.**

Where are we left? We are left with the prospect of having to reject the standard methodological practice in epistemology. The goal of epistemology is to tell us what makes justified beliefs justified. The standard practice is to check to see which epistemological theory is in reflective equilibrium with our shared and nonnegotiable epistemological intuitions. But this practice is ill-suited to this goal, since intuition sorely underdetermines theory choice in epistemology, that is, since many of the major theories of justification are in reflective equilibrium with the relevant class of intuitions. Since there is apparently no way to shore up the reflective equilibrium test to ameliorate this problem, the standard practice should be rejected.

Earlier in the second section, I provided some initial motivation for thinking that the 'intuitional adequacy' of divergent epistemological theories should have been expected. That was to point out that some of these theories have been tested again and again against intuition and refined accordingly over the course of decades. Now I want to give some further motivation for thinking that an underdetermination problem should have been expected. Rawls himself, for instance, seems to confront the problem head on as far back as A theory of justice:

[D]oes a reflective equilibrium (in the sense of the philosophical ideal) exist? If so, is it unique? Even if it is unique, can it be reached? ... I shall take for granted that these principles [i.e., principles that characterize our intuitions] are either approximately the same for persons whose judgments are in reflective equilibrium, or if not, that their judgments



divide along a few main lines represented by the family of traditional doctrines that I shall discuss. (1971 50, emphasis mine)

The question whether there is a unique and attainable state of reflective equilibrium is indeed an interesting one. It is also a broadly empirical one. Rawls rightly notes that it would carry him beyond the scope of A theory of justice to address it. But that is not the problem. The problem is that Rawls simply assumes, as if it does not matter which, that either our moral intuitions will yield more or less the same system of principles in reflective equilibrium or they will “divide along a few main lines” of traditional moral theories. But in the latter case, where our intuitions differ, it is hard to see how Rawls can go on to argue for his own view over its rivals by the method of reflective equilibrium. It is not as though Rawls is not interested in converting his opponents. Among those other traditional moral theories is utilitarianism, a view Rawls is very much interested in clearing out of the picture. Now if intuitions divide along two lines, say utilitarianism and Rawlsian justice-as-fairness, then Rawls cannot argue from shared intuitions (i.e., intuitions common to utilitarians and Rawlsians), via reflective equilibrium, to justice-as-fairness. Shared intuition would underdetermine a choice between utilitarianism and justice-as-fairness.

So it looks like we should have anticipated our underdetermination problem long ago. And since this point is made in connection with Rawls, one naturally suspects that there may be an exactly parallel underdetermination problem in ethics. Do Kantians negotiate down on their consequentialist intuitions to reach a point of reflective equilibrium? Do utilitarians negotiate down on their deontologist intuitions to strike an agreement between principles and particular judgments? It would be dangerous to conclude the matter too quickly here; but normative epistemology and normative ethics are very similar metaphysical projects, and insofar as they employ the same methodology, it is not unreasonable to suspect that they may have similar methodological problems.

## REFERENCES

- Ackermann, Ruby and Robert J. DeRubeis. 1991. “Is ‘depressive realism’ real?” Clinical Psychology Review 5: 565-84.
- BonJour, Laurence. 2002. Epistemology: classic problems and contemporary responses. (Lanham, MD: Rowman & Littlefield).
- Conee, Earl. 1996. “Plantinga’s naturalism,” in Kvanvig (1996): 183-96.
- Daniels, Norman. 1979. “Wide reflective equilibrium and theory acceptance in ethics” Journal of Philosophy 76: 256-82.
- Dennett, Daniel C. 1978. “Intentional systems” in Brainstorms. (Cambridge, MA: MIT Press).

- DePaul, Michael R. and William Ramsey (eds). 1998. Rethinking intuition. (Lanham, MD: Rowman & Littlefield).
- Goldman, Alvin. 1979. "What is justified belief?" in Justification and knowledge, ed. by George Pappas. (Dordrecht: Reidel): 1-23.
- \_\_\_\_\_. 1986. Epistemology and cognition. (Cambridge, MA: Harvard University Press).
- \_\_\_\_\_ and Joel Pust. 1998. "Philosophical theory and intuitional evidence" in DePaul & Ramsey (1998): 179-97.
- Goodman, Nelson. 1965. Fact, fiction and forecast, 2<sup>nd</sup> ed. (Indianapolis: Bobbs-Merrill).
- Graham, George and Terry Horgan. 1998. "Southern fundamentalism and the end of philosophy" reprinted in DePaul & Ramsey (1998): 271-92.
- Hempel, Carl G. 1966. The philosophy of natural science. (Englewood Cliffs, NJ: Prentice-Hall).
- Jackson, Frank. 1998. From metaphysics to ethics: a defense of conceptual analysis. (Oxford: Clarendon Press).
- Kim, Jaegwon. 1988. "What is 'naturalized epistemology?'" Philosophical Perspectives 2: 381-405.
- Kvanvig, Jonathan L. 1996. Warrant in contemporary epistemology. (Lanham, MD: Rowman & Littlefield).
- Lehrer, Keith. 1990. Theory of knowledge. (Boulder, CO: Westview Press).
- \_\_\_\_\_. 1996. "Proper function vs. systematic coherence" in Kvanvig (1996): 25-45.
- Markie, Peter. 1997. "In defense of one form of traditional epistemology." Philosophical Studies 85: 37-55.
- Plantinga, Alvin. 1993a. Warrant: the current debate. (New York: Oxford University Press).
- \_\_\_\_\_. 1993b. Warrant and proper function. (New York: Oxford University Press).
- Quine, W.V.O. 1953. "Two dogmas of empiricism" in his From a logical point of view. (Cambridge, MA: Harvard University Press).
- \_\_\_\_\_. 1969. "Epistemology naturalized" in his Ontological relativity and other essays. (New York: Columbia University Press).
- Rawls, John. 1971. A theory of justice. (Cambridge, MA: Harvard University Press).