Is Reflective Equilibrium Enough?

Thomas Kelly
Princeton University

Sarah McGrath
Princeton University

1. Introduction

Suppose that one is at least a *minimal realist* about a given domain, in that one thinks that that domain contains truths that are not in any interesting sense of our own making. Given such an understanding, what can be said for and against the *method of reflective equilibrium* as a procedure for investigating the domain?

One fact that lends this question some interest is that many philosophers do combine commitments to minimal realism and a reflective equilibrium methodology. Here, for example, is David Lewis on *philosophy*:

> Our "intuitions" are simply opinions: our philosophical theories are the same. Some are commonsensical, some are sophisticated; some are particular; some general; some are more firmly held, some less. But they are all opinions, and a reasonable goal for a philosopher is to bring them into equilibrium. Our common task it to find out what equilibria there are that can withstand examination, but it remains for each of us to come to rest at one or another of them…
>
> Once the menu of well-worked out theories is before us, philosophy is a matter of opinion. Is that to say that there is no truth to be had? Or that the truth is of our own making, and different ones of us can make it differently? Not at all! If you say flatly that there is no god, and I say that there are countless gods but none of them are our worldmates, then it may be that neither of us is making any mistake of method. We may each be bringing our opinions to equilibrium in the most careful possible way, taking account of all the arguments, distinctions, and counterexamples. But one of us, at least, is making a mistake of fact. Which one is wrong depends on what there is (1983: x-xi).

In addition to philosophy in general, the method of reflective equilibrium has also been endorsed as the appropriate procedure for investigating various other subject matters, including logic and inductive reasoning (Goodman 1953), and especially, normative ethics and political philosophy.[1] Indeed, prominent moral philosophers sometimes suggest that when it comes to moral inquiry, the method of reflective equilibrium is, in effect, the only game in town. Thus, according to Michael Smith, it is among the "platitudes" about morality that properly conducted moral inquiry has "a certain characteristic coherentist form", of a kind that was given systematic articulation by John Rawls (Smith 1994: 40-41). Similarly, according to Thomas Scanlon:

> …it seems to me that this method, properly understood, is in fact the best way of making up one's mind about moral matters and about many other subjects. Indeed, it is the only defensible method: apparent alternatives to it are illusory (2002:149).[2]

Nevertheless, the method of reflective equilibrium has been fiercely criticized since its earliest explicit formulations.[3] A common charge among detractors is that the method is *too weak*, in the following respect: even if one impeccably executes the method, the views at which one arrives might nevertheless be hopelessly inadequate. Many of the more specific charges brought against the method—for example, that it is overly conservative, in the sense that it unduly privileges the beliefs that one holds before inquiry begins—can be seen as variations on this more general theme.

Notice, however, that if there is some compelling objection along these lines, the charge cannot simply be that impeccably executing the method could fail to lead us to the

[1] See, e.g., Daniels (1996, 2003), DePaul (1998, 2006), Harman (2004), McMahan (2000), Rawls (1971, 1993, 1999, 2001), Scanlon (2002), and Smith (1994), among many others.

[2] Compare DePaul (2006: 616) who argues that, when it comes to moral inquiry, "there is simply no reasonable alternative to reflective equilibrium".

[3] Important early critics include Hare (1973), Singer (1974), Lyons (1975), and Brandt (1979, 1990); prominent later critics include Copp (1985), Cummins (1998), and Stitch (1990).

truth, or even that doing so could lead us to views that are radically mistaken. For no clear-headed realist should accept the idea that it is a condition of adequacy on a method of inquiry that that method is guaranteed to deliver the truth, or even that it will not leave us much worse off with respect to the truth than if we had never availed ourselves of it. Certainly, we do not hold our best scientific methods to the relevant standard. In a world in which the empirical evidence that we have to go on is consistently misleading or unrepresentative—either because of the chicanery of an evil demon, or through simple long-run bad luck—the impeccable application of our best scientific methods will not only fail to deliver the truth but will lead us further and further astray. No realist should think that this is a good objection to those methods. Similarly, it is not a good objection to the method of reflective equilibrium that there are circumstances in which employing it could lead us into error, even radical error.[4]

Thus, the charge that the method is too weak must be understood in some other way. For example, we believe that it *would* be a good objection to the method if it turned out

[4] For this reason, charges that (e.g.) the method is overly conservative must be put with some care if they are not to miss the mark entirely. Again, the charge cannot simply be that, if the beliefs from which we begin are sufficiently mistaken, then even perfect application of the method will fail to lead us to the truth. That much is surely plausible, but it is dubious that any plausible methodology lacks the feature in question. Indeed, we think that one should be positively suspicious of any account of methodology that is advertised as lacking the feature in question. The discovery of interesting truths about normative ethics or politics (or truths of philosophical ontology, etc.) is, one suspects, no mean feat even in relatively favorable circumstances. A case in which our pre-philosophical views about what is morally required of us or what exists are radically in error is a case in which we are maximally ill-positioned to discover such truths. It is one in which we sit down to play an exceedingly difficult game having been dealt a particularly bad hand. If these were our circumstances, it would be a mistake to assume that a good method would provide us with a rational path out of the darkness and into the light.

This is not to say that there is no cogent objection to the method on the grounds that it is overly conservative, only that the charge of conservatism must be developed with greater care than is sometimes done, if it is to have a chance of being cogent.

that impeccably executing it could lead one to hold views that are *unreasonable* for one to hold. (And no doubt, this is what many of its critics have had in mind.) For surely, if some method is in fact the best method for investigating some domain, and one employs the method because one recognizes that this is so, then the views at which one arrives by impeccably executing it would not be unreasonable. Thus, if one could arrive at unreasonable views by impeccably executing the method of reflective equilibrium, it follows that it is not the best method.

One might think that requiring that the method of reflective equilibrium not lead to unreasonable beliefs is too stringent, for reasons analogous to those that speak against a requirement that the method not lead to false beliefs. For imagine an individual who begins with views about (say) morality that are completely unreasonable. Suppose that the individual pursues and achieves a state of reflective equilibrium by reasoning flawlessly "downstream" from that rationally defective starting point. If the views at which the person arrives are intuitively unreasonable, then one might suggest that this should not be held against the method, for the method cannot be expected to deliver reasonable outputs given unreasonable inputs. On this account, the goodness of the method of reflective equilibrium as a procedure would be something like the goodness of reasoning in accordance with modus ponens. If one reasons from two unreasonable beliefs to a third belief in accordance with modus ponens, then the third belief might very well be unreasonable as well, but surely this is not a good objection to the practice of reasoning in accordance with modus ponens. Similarly, one might think, it is too much to require that the method of reflective equilibrium not lead to unreasonable beliefs when a person begins from a rationally defective starting point.

This picture sets the bar too low. Although natural, we do not believe that such comparisons do justice to the role that proponents of the method of reflective typically claim for it. Proponents of the method typically claim that it is *the* appropriate method for investigating this or that domain; it is not simply one norm or rule among many others (e.g., "One should seek coherence among one's views") which is what the comparison with modus ponens suggests. After all, someone who thinks that the method of reflective equilibrium is hopelessly inadequate as a characterization of correct methodology in ethics might very well agree that one should seek coherence among one's moral beliefs. (Consider, for example, a philosopher who thinks that our ability to arrive at moral knowledge depends essentially on the operation of an occult, *sui generis* faculty of moral intuition, and that no account of moral methodology that fails to mention the central role of this faculty could possibly be adequate.) In this respect, the method of reflective equilibrium purports to play the same role as the cluster of procedures that are employed by (e.g.) physicists and biologists in investigating their respective domains.

Suppose that, prior to embarking upon the systematic study of fruit flies, one held various baseless opinions about their nature. If one then devoted oneself to the study of fruit flies, and impeccably followed the best scientific procedures we have for arriving at accurate views about their nature, we would expect those earlier baseless opinions to be filtered out or corrected at some stage in the inquiry. In the unlikely event that some of those opinions were among the views that one held after having impeccably following our best scientific methods, then, we submit, those beliefs would no longer be unreasonable ones to hold. If someone *did* criticize them as unreasonable, one would be in a position to reply as follows:

> My views about fruit flies are ones that have withstood the impeccable application of our best methods for arriving at and correcting beliefs about fruit flies. Therefore, whatever else is true of these beliefs (e.g., even if later inquiry should show that they are false), they are not unreasonable views for me to hold as things stand.

We think that this would be an excellent defense. Similarly, if the method of reflective equilibrium really is the best method for arriving at one's views in some domain, then it would be a good defense of the reasonableness of those views that they either resulted from or withstood the impeccable application of that method. And therefore, it would be a good objection to the method if it were shown that one could arrive at unreasonable beliefs by employing it.[5]

In point of fact, proponents of the method typically think that there are significant constraints on admissible starting points: thus, if one simply sets out from all of one's initial opinions, no matter how baseless or ill-considered, then one is not competently applying the method. (In the broadly Rawlsian tradition, this is the idea that the correct starting point consists of our *considered judgments*.) We will consider this idea at some length below.

In addition to the worry that the method licenses unreasonable beliefs, there are other ways in which the charge that it is too weak might be developed. For example, in the passage quoted above, Lewis suggests that two philosophers might competently execute

[5] In fact, the argument of the preceding paragraphs oversimplifies things in one respect. That a given method is the best method for investigating a given domain (and is known to be so) is not strictly speaking a sufficient condition for the reasonableness of the views to which it leads. For suppose that we had *no* good methods for investigating a given domain: even our best method is highly unreliable, and known to be so. In that case, it would not be a good defense of the reasonableness of some belief to show that it was sanctioned by the best method. But of course, proponents of the method of reflective equilibrium typically do not think that it is a poor method that nevertheless manages to be the best of a bad lot.

the method and yet arrive at very different equilibria, even if they both take into account all of the same arguments, distinctions, and counterexamples. Although Lewis apparently did not regard this putative possibility as a reason to doubt the method, one might plausibly hold that a good method should lead rational inquirers to converge in their views, at least if they are exposed to the same considerations. (Notice that this concern is independent of the previous one, inasmuch as one who is moved by it need not hold that inquirers who settle on different equilibria are unreasonable for believing as they do.)

In what follows, we will explore the idea that the method of reflective equilibrium is too weak in greater detail. Thus far, our discussion has been relentlessly abstract; in order to anchor it, we will critically examine the accounts of the procedure offered by three of of its most influential and philosophically sophisticated proponents. We will begin with the seminal accounts of Nelson Goodman (1953) and John Rawls (1971), and then turn to the more recent discussion of Thomas Scanlon (2002).

2. Goodman and Coherence

Remarkably, Goodman's "The New Riddle of Induction" stands as a classic of twentieth century philosophy for two independent reasons. Undoubtedly, it is most famous for introducing the philosophical problem that gives the essay its name. Our concern, however, is with Goodman's discussion of what he called "the old problem of induction"—that is, the kind of skepticism about inductive reasoning associated with David Hume. For in the course of attempting to "dissolve" Humean skepticism about induction, Goodman offered arguably the first clear statement of what Rawls would later dub "the method of reflective equilibrium". The crucial passage is worth quoting at some length:

> How do we justify a *de*duction? Plainly, by showing that it conforms to the general rules of deductive inference…Analogously, the basic task in justifying an inductive inference is to show that it corresponds to the general rules of induction…
>
> The validity of a deduction depends not upon conformity to any purely arbitrary rules we may contrive, but upon conformity to valid rules…But how is the validity of the rules to be determined? Here we encounter philosophers who insist that these rules follow from some self-evident axiom, and others who try to show that the rules are grounded in the very nature of the human mind. I think the answer lies much nearer the surface. Principles of deductive inference are justified by their conformity with accepted deductive practice. Their validity depends upon accordance with the particular deductive inferences we actually make and sanction. If a rule yields unacceptable inferences, we drop it as invalid. Justification of general rules thus derives from judgments rejecting or accepting particular deductive inferences.
>
> This looks flagrantly circular. I have said that deductive inferences are justified by their conformity to valid general rules, and that general rules are justified by their conformity to valid inferences. But this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. *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.* The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either (63-64, emphasis his).

Having sketched this general picture of justification with respect to deduction, Goodman then applies it, *mutatis mutandis*, to the case of induction. Thus, we justify particular inductive inferences by showing that they correspond to principles of induction that we actually accept, and those principles are justified in turn by showing that they correspond to our judgments about which particular inferences are acceptable and which are unacceptable. In this way, Goodman claims, Humean skepticism about induction is effectively dissolved.

Suppose that one infers:

The bread that has always nourished me in the past will do so again today.

On Goodman's account, justifying this particular inference is a matter of showing that it conforms to accepted inductive practice, i.e., that it is sanctioned by some inductive principle that we actually accept. Let us say that the corresponding belief is *Goodman-justified* just in case this condition is met. Notably, even a full-fledged inductive skeptic, i.e., someone who flatly denies that we have any inductive knowledge at all, will allow that this belief is Goodman-justified. After all, the inductive skeptic does not deny that the relevant inference is in accordance with our actual inductive practice; rather, he denies that its being in accordance with that practice is of any epistemic significance, in light of the considerations adduced by Hume. He sees no reason to think that beliefs about the future that are Goodman-justified are more likely to be true, or better candidates for knowledge, than beliefs that are not Goodman-justified. Thus, the fact that some of our beliefs are Goodman-justified, and even facts about *which* beliefs are Goodman-justified, would seem to be undisputed common ground between the inductive skeptic and the non-skeptic. Given this, one might doubt whether anything that Goodman says about justification in this context tells even slightly in favor of the non-skeptic as against the skeptic. Indeed, one might very well wonder: how could Goodman himself have thought otherwise?

The short answer to the last question is: He didn't. Although the point is not often emphasized, Goodman himself seems to have been a full-fledged inductive skeptic at the time he wrote "The New Riddle of Induction". As evidence of this, consider the following passage, in which Goodman is giving his view about what "Hume's problem" is *not*:

> If the problem is to explain how we know that certain predictions will turn out to be correct, the sufficient answer is that we don't know any such thing. If the problem is to find some way of distinguishing antecedently between true and false predictions, we are asking for prevision rather than for philosophical explanation. Nor does it help matters much to say that we are merely trying to show that or why certain predictions are probable…obviously the genuine problem cannot be one of attaining unattainable knowledge or of accounting for knowledge that we do not in fact have…(p.62).

Consider the following two inconsistent predictions:

(1) Of the human beings alive today, some will not be alive in fifty years time.

(2) Of the human beings alive today, all will still be alive in fifty years time.

According to the view articulated by Goodman, we do not know which of these two predictions will turn out to be correct, and we lack any way of distinguishing the true prediction from the false prediction. Clearly, this is a radical claim. Indeed, we believe that this passage from Goodman is as explicit an endorsement of distinctively inductive skepticism as one finds in the history of philosophy. (Certainly, it is at least as clear an endorsement as anything that one finds in Hume himself.)

Significantly, Goodman's disavowal of genuine inductive knowledge occurs immediately before he describes the reflective equilibrium conception of justification. We think that this is no accident, and that Goodman's attempt to *deflate the explanandum* ("obviously the genuine problem cannot be one of attaining unattainable knowledge, or of accounting for knowledge that we do not in fact have") plays a key role in his overall argument. On a traditional conception of justification, a belief is justified just in case it would amount to knowledge provided that it is true.[6] Thus, to say that we are justified in believing that *not everyone alive today will still be alive in fifty years time*, is to say that

[6] Of course, since Gettier (1963), it has generally been thought that justified true belief is insufficient for knowledge. We do not believe that this makes a material difference to the points which follow, so we will ignore complications created by Gettier cases.

our basis for thinking that this proposition is true is sufficiently strong that our belief qualifies as knowledge provided that it is true. Offhand, however, Goodman-justification looks too weak to underwrite genuine knowledge. After all, in principle, there is nothing that precludes the possibility that an inductive principle that passes all of Goodman's tests with flying colors is in fact highly unreliable. (We do not believe that Goodman would have disagreed with this.) In that case, the inductive conclusions sanctioned by this principle are Goodman justified, despite the fact that the vast majority of them are false. Given this, it seems that even those relatively few conclusions that are true fail to count as known, in view of the general unreliability of the principle. So Goodman justification seems like a poor candidate for justification in the traditional sense of that which underwrites knowledge.

Of course, from Goodman's perspective this is no objection to his account of justification, for we are not in a position to have inductive knowledge: at least with respect to our beliefs about the future, justification in any stronger sense is chimerical. In effect, in disavowing inductive knowledge, Goodman is disavowing any pretense that Goodman justification amounts to justification in the traditional sense of that which underwrites knowledge. For Goodman, a *solution* to Hume's problem would—if such a thing were possible—show how inductive knowledge is possible, or at least that certain inductive conclusions are known. But for exactly this reason, Goodman explicitly disavows any claim to having solved Hume's problem; rather, he has dissolved Hume's problem by showing that a widespread conception of it rests on a false presupposition (viz. that we have inductive knowledge). It is only once the explanandum has been thus deflated—in showing how some inductive inferences can be justified, we are not vindicating the possibility of inductive knowledge--that the conception of justification on offer ceases to look vulnerable to what would otherwise be an obvious objection, viz. that Goodman justification is too weak to underwrite knowledge of the future.

Pace Goodman, however, inductive skepticism is false. For example, here are a few things that we know about the future:

(1) Not everyone who is currently alive will still be alive fifty years from now.
(2) Some of the people who are currently alive will still be alive ten seconds from now, and
(3) Some of the people who are currently alive will not die of leukemia.

As we have seen, Goodman thought that the fact that a true belief about the future is justified in his sense does not mean that it is knowledge. For the reason given above, we believe that he was right about this: the mere fact that a given belief about the future is both true and held in a state of reflective equilibrium does not mean that it is knowledge, since its satisfying the relevant conditions is consistent with its being the deliverance of a highly unreliable inductive principle. However, given that we *do* have at least some knowledge of the future, it follows immediately that there is some *other* epistemological story to be told about such knowledge: our knowledge of the future is not (simply) a matter of the fact that some of our beliefs about the future are both true and held in a state of reflective equilibrium.

Before taking leave of Goodman, we should note an aspect of his account of justification that contributes to the sense that justification so understood is too weak to underwrite knowledge. Recall Goodman's claim that

> The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either.

The idea that 'in the agreement achieved lies the *only* justification needed for either' is characteristic of a *coherentist* as opposed to a foundationalist account of justification. For any reasonably sophisticated foundationalist will admit that considerations of coherence can contribute to (or detract from) the epistemic status of one's beliefs; what the foundationalist will adamantly deny is that coherence could be the *entire* story about justification. Typically, the foundationalist will insist that at least some beliefs ('properly basic' or foundational beliefs) enjoy at least some measure of rational credibility or positive epistemic status apart from considerations of coherence, and that, if this were not so, *no* beliefs would be justified, no matter how well-integrated they are within a coherent set. In contrast, it is characteristic of the coherentist to insist that an adequate level of coherence is *sufficient* for justification, and it is this characteristic commitment to which Goodman signals his allegiance here.

In fact, the dominant understanding of the method of reflective equilibrium seems to be one on which it is a kind of dynamic coherence theory.[7] So understood, the method of reflective equilibrium invites all of the standard objections that are raised for coherentist accounts of justification. In the passage in which he describes the method, Goodman alludes to one such standard objection, viz. that the envisaged justification is *circular*. In response, he offers a standard coherentist reply--that the circularity in question is virtuous, not vicious. More relevant for our purposes, however, is another classic concern about coherence theories: doubts about whether the mere coherence of a belief system could ever underwrite knowledge or even justified beliefs about an independent subject

[7] On this point, see especially Norman Daniels' survey (2003). Explicit exceptions to the tendency to interpret the method in coherentist terms include Harman's (2004) 'general foundations' interpretation of the method and McMahan (2000).

matter. After all, how coherent a system of beliefs is is, presumably, something that supervenes on the relations that obtain between those beliefs, as opposed to any relations that obtain between those beliefs and anything outside the system. But this makes salient the possibility that a system of beliefs could be arbitrarily coherent while being radically detached from the very subject matter that it purports to accurately represent. To be clear, the problem is not that a coherentist account allows for the possibility that a highly justified set of beliefs could be more or less entirely in error. Indeed, as noted above, it is plausible that allowing for this possibility is a desideratum (if not an outright condition of adequacy) for any account of justification, since, intuitively, an individual in sufficiently unfortunate circumstances might have a radically false view of things despite having beliefs that are highly justified (Cohen 1984). Rather, the problem is that, at least in principle, an individual might maintain a perfectly coherent set of beliefs while being completely unresponsive to relevant and easily perceptible changes in his or her environment. This is the point exploited by stock counterexamples to coherentism about justification in the epistemological literature.[8] Intuitively, an individual who simply maintained the same perfectly coherent set of beliefs about her environment, despite the fact that her experiences of that environment were constantly changing, would not be justified in holding those beliefs.

In light of this 'No Contact with Reality' objection, coherentist theories of justification have always looked particularly implausible when offered as accounts of that which underwrites empirical knowledge. Indeed, prominent twentieth century

[8] See, e.g., Feldman's (2003: 68) 'Strange Case of Magic Feldman' and Plantinga's (1993: 82) 'Case of the Epistemically Inflexible Climber'.

philosophers who embraced coherentist accounts of empirical knowledge were sometimes led to idealism (Blanshard 1939) or coherentist accounts of truth (Hempel 1934-35a,b) in an attempt to bridge the gap.[9] Similarly, the method of reflective equilibrium, when understood as a dynamic coherence theory, does not seem particularly plausible as an account of how empirical scientists should arrive at their views of how the world works, given that it makes no essential reference to observation or perception.

Suppose that that much is conceded, and consider two different (though compatible) responses that a proponent of the method might offer. First, she might restrict the domains for which the procedure is claimed to provide an appropriate methodology. For example, even if the procedure would be an inappropriate methodology for investigating empirical matters of fact, it does not follow that it is an inappropriate methodology for investigating normative ethics, political philosophy, or philosophy more generally. After all, many of strongest objections to global coherentist accounts trade on the apparent inability of such accounts to do justice to the role of experience, or empirical observation. But of course, counterexamples of the relevant kind will not be available in domains where inquiry is not driven by empirical observation. Secondly, a proponent of the method might attempt to understand it, not as a coherence theory, but rather as a kind of foundationalism, albeit a variety in which considerations of coherence play a large role. We will consider instances of both of these strategies in what follows.

3. Rawls and Convergence

[9] In recent years, BonJour (1985) is arguably the most ambitious and widely discussed attempt to show how a coherentist account of empirical justification can be combined with a realist conception of truth. In his contribution to BonJour and Sosa (2003), he abandons the project as unworkable and advocates a return to a relatively traditional form of foundationalism.

The fact that so many contemporary philosophers explicitly conceive of their own methodology in terms of the reflective equilibrium picture surely owes more to the influence of Rawls than any other individual. More specifically, the widespread popularity of that conception of methodology among moral and political philosophers is due in large part to Rawls' championing of the method in *A Theory of Justice* (1971). Although there are important differences that we will explore, in broad outline Rawls' account of the method in the moral and political domain is similar to the account that Goodman gives in the context of discussing deduction and induction. Here is the account that Rawls offers in "The Independence of Moral Theory" (1974):

> People have considered judgments [about morality] at all levels of generality, from those about particular situations and institutions up through broad standards and first principles to formal and abstract conditions on moral conceptions. One tries to see how people would fit their various convictions into one coherent scheme, each considered judgment whatever its level having a certain initial credibility. By dropping and revising some, by reformulating and expanding others, one supposes that a systematic organization can be found. Although in order to get started various judgments are viewed as firm enough to be taken provisionally as fixed points, there are no judgments on any level of generality that are in principle immune to revision (p.289).

By proceeding in this way, one attempts to bring one's moral convictions into a state of reflective equilibrium. Crucially, for Rawls the state that we should pursue is one of *wide* (as opposed to 'narrow') reflective equilibrium. The pursuit of wide reflective equilibrium is the pursuit of a comprehensive moral view that "would survive the rational consideration of all feasible moral conceptions and all reasonable arguments for them" (1974: 289).[10] Of course, Rawls acknowledges that it is not realistic that we will actually

[10] Although the terminology of 'wide' reflective equilibrium is introduced in later work, the idea is clearly present in *A Theory of Justice*. There, Rawls writes of a state of equilibrium that is reached after having considered "all possible descriptions to which

consider all such conceptions and arguments.[11] Rather, for Rawls, the state of wide reflective equilibrium constitutes an ideal: it is the hypothetical end point of properly conducted moral inquiry, if such inquiry were pursued without limit.

In addition to the idea of wide reflective equilibrium, a second significant innovation introduced by Rawls is the apparatus of *considered judgments* as that on which the process of seeking reflective equilibrium operates. For Rawls, "considered judgment" is a technical term.[12] Not everything that one believes or judges true, even on reflection, qualifies as a considered judgment. Rather, considered judgments are judgments of which one is confident (as opposed to uncertain or hesitant), that are issued when one is able to concentrate without distraction on the question at hand (as opposed to when one is 'upset or frightened') and with respect to which one does not stand to gain or lose depending on how the question is answered. In addition, such judgments must be stable over time.

Of course, the point behind the introduction of considered judgments is that "in deciding which of our judgments to take into account, we may reasonably select some

one might plausibly conform one's judgments together with all relevant philosophical arguments for them" (1971:49).

[11] See (1971:49) and, more definitively, (1974:289). Cf. Scanlon (2002:141): "It should be emphasized that this is not a state that Rawls believes we are currently in, or likely to reach. It is rather an ideal, the struggle to attain which continues indefinitely".

[12] And indeed, one whose stipulated meaning changed considerably from work to work. For example, in the early "Outline of a Decision Procedure for Ethics" (1951) a considered judgment must concern actual (as opposed to merely hypothetical) cases (p.5), and cannot be the object of disagreement among 'competent persons' (p.6); both of these requirements are absent from later characterizations. In *A Theory of Justice*, considered judgments concern particular cases; in that work, "considered judgment" is frequently juxtaposed with "general conviction" or "general principle". As the above passage from "The Independence of Moral Theory" makes clear, however, by the time of that work Rawls was applying the term to judgments of all levels of generality. In what follows, we will work with this last and most general formulation.

and exclude others" (1971: 47). Thus, for Rawls, there are at least two different ways in which a moral conviction can be legitimately discarded: (i) it might fail to qualify as a considered judgment, or (ii) it might qualify as a considered judgment, but be eliminated at some later stage in the course of pursuing reflective equilibrium. Because many moral judgments might fail to qualify as considered judgments, a significant amount of filtering might occur even before the process of seeking reflective equilibrium begins. Significantly, although Rawls is often read as a coherentist, this last fact opens the door to the possibility of putting a more foundationalist spin on his account. Presumably, a moral belief that qualifies as a considered judgment has some positive epistemic status that is not had by those beliefs that fail to qualify as such; moreover, that positive epistemic status is not exclusively a matter of its cohering well with the rest of what one believes. (And indeed, notice that in the passage quoted above, Rawls speaks of considered judgments as each having 'a certain level of initial credibility'.) In fact, it seems that the following kind of modest foundationalism is consistent with Rawls' general framework: any considered judgment is immediately justified, i.e., justified in a way that is not a matter of the relations that it stands in to other beliefs. This justification is defeasible, however, and it is defeated if the considered judgment cannot be made to adequately cohere with the rest of what one believes.[13]

Although it seems to be consistent with his general framework, we do not attribute this view to Rawls. Indeed, we do not believe that the relevant texts warrant attributing to Rawls a general view about the conditions under which a particular moral belief or

[13] The resulting view would be quite close to the 'general foundations' theory championed by Harman (2004).

judgment is justified for an individual. Perhaps it is safe to take the following as a sufficient condition:

> A moral judgment is justified for an individual if she holds it in a state of wide reflective equilibrium.

Notice, however, that this sufficient condition rarely if ever obtains, inasmuch as wide reflective equilibrium constitutes an ideal that is rarely if ever achieved. Presumably, however, some of our current moral beliefs are justified even if we are not currently in a state of wide reflective equilibrium. Let us set this issue aside, however, and return to questions about the suitability of the method to achieving the goals of inquiry.

It is natural to think that knowledge is a goal of inquiry (perhaps even *the* goal of inquiry), and that a good method for investigating a domain is one that is well-suited to deliver knowledge of that domain, or at least, more likely than whatever alternative methods might be available. Even if one thinks that full-fledged knowledge is off the table (as Goodman thought in the case of our beliefs about the future), one might still take *truth* as the goal of inquiry, and evaluate one's methods in terms of their suitability for achieving that goal.[14] Construed along these lines, the goal of moral philosophy would be that of arriving at the truth about what is right or wrong, what we are morally required to do, and so on. Questions about the potential strengths and weaknesses of the

[14] If one thinks that knowledge in some domain is off the table, shouldn't one also be skeptical about one's ability to evaluate methods in terms of their ability to arrive at the truth? Not necessarily, especially if one is involved in making comparative evaluations among methods. For example, Reichenbach (1938) thought that Hume's critique of inductive reasoning suffices to show that we are not in a position to have either inductive knowledge or knowledge that our actual inductive methods are reliable; nevertheless, he argued that those methods weakly dominated any other method that we might employ with respect to arriving at true beliefs about the future.

method of reflective equilibrium would thus be questions about its suitability as a means for achieving this goal.

Interestingly, this is not how Rawls generally thinks about the aims of moral philosophy. In *A Theory of Justice* (46), he provisionally characterizes moral philosophy as the attempt to describe our underlying "moral capacity" or "moral sensibility" (or, in the distinctively political sphere, "our sense of justice"). Elsewhere, he says that the aim of the method of reflective equilibrium is to investigate the underlying "substantive moral conceptions" that people actually hold; the procedure is thus "a kind of psychology, and does not presuppose the existence of objective moral truths" (1999: 290). This orientation seems to be largely motivated by Rawls' belief that "the history of moral philosophy shows that the notion of moral truth is problematical" (1999: 290). Significantly, in "The Independence of Moral Theory" (1974), perhaps Rawls' most explicitly methodological essay, it is only *after* the possibility that there are moral truths has been bracketed or provisionally set aside that the method of reflective equilibrium is brought on stage and described; it is then touted as that procedure best suited to achieving the descriptive, psychological task of uncovering substantive moral conceptions.[15]

In his interpretation of Rawls on reflective equilibrium, Scanlon (2002) distinguishes between two interpretations of the method. On the *deliberative* interpretation, the aim of the method is to determine what to believe about morality or justice. On the *descriptive* interpretation, the aim of the method is to describe the underlying moral conception or

[15] For further denigration of the idea that moral truth is the proper aim of moral inquiry, see also his (1980): 306-307.

sense of justice that is held by a particular person (perhaps oneself) or group of people.[16] Although a great deal of what Rawls says about reflective equilibrium suggests the descriptive interpretation, let us set it aside and concentrate on the deliberative interpretation, on which it is a procedure for figuring out what to believe, or the truth about morality. What can be said for and against the method as a tool for achieving this goal? One might think that a good method for investigating a given domain would have the following property: if the method is impeccably employed by different individuals, then those individuals would tend to converge in their views over time, at least if they were exposed to the same considerations. Rawls himself was much concerned with questions about whether the method of reflective equilibrium would lead to a convergence among those who employed it. In *A Theory of Justice*, he raised, but did not pursue, the following issues:

> This explanation of reflective equilibrium suggests straightaway a number of further questions. For example, does 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? Perhaps the judgments from which we begin, or the course of reflection itself (or both), affect the resting point, if any, that we eventually achieve (p.50).

Consider the issue of whether there is a unique reflective equilibrium. Presumably, there are at least two questions here:

[16] As Scanlon notes, the rationale for certain aspects of the method will differ depending on what interpretation is in play. Consider, for example, the fact that only considered judgments are to be taken into account. On the deliberative interpretation, this restriction is motivated by the fact that considered judgments are (presumably) more likely to be true judgments about morality or justice than judgments that fail to qualify as such. On the descriptive interpretation, the restriction is motivated by the thought that considered judgments more accurately reflect the underlying conception of the person whose moral sensibility is being described.

(1) The intrapersonal question: for any particular person, is there some unique reflective equilibrium that she would arrive at if she employed the method impeccably?

(2) The interpersonal question: would different individuals, each of whom employed the method impeccably, converge on a unique reflective equilibrium?

Consider first question (1). Given that one's considered moral judgments are currently not in equilibrium, is there any reason to suppose that there is some rationally optimal way for one to resolve those conflicts that exist? Offhand, it seems that there might be multiple ways of achieving perfect coherence, resulting in at least somewhat (and perhaps even radically) different sets of judgments. Of course, what is relevant here is *wide* reflective equilibrium. Perhaps if one were presented with "all feasible moral conceptions and all reasonable arguments for them", one would be rationally compelled to resolve those conflicts in exactly one way, and be driven to some specific equilibrium. Although it is far from obvious, let us simply assume that this is how things would transpire; more generally, let us assume for the sake of argument that the answer to question (1) is 'Yes'.

Still, it does not follow that for different individuals there is a unique reflective equilibrium. In general, that (1) receives an affirmative answer is a necessary but insufficient condition for (2)'s receiving an affirmative answer. If the answer to (1) is affirmative, then, for any particular set of initial considered judgments that a person might hold, there is some unique reflective equilibrium that would be reached by impeccably applying the revision procedure to that set. Even if that is true, it of course does not follow that impeccably applying the procedure to a different set of initial starting points would lead to the same state. Indeed, at least offhand, this seems rather unlikely. Perhaps the following is among one's considered moral judgments:

> Even if a doctor could save the lives of two people dying for want of some vital organ by forcibly overpowering and harvesting the organs of some innocent and unwilling bystander, it is morally impermissible for her to do so.

If so, then in all likelihood, one also holds other considered judgments with which this judgments coheres. Someone with act utilitarian sympathies might have, among his considered judgments, the judgment that in the envisaged scenario the doctor is not only permitted but *morally required* to harvest the organs of the bystander; no doubt, that judgment coheres well with other things that he believes. Given these radical differences, why think that the best way for each person to achieve coherence among his or her *own* judgments will lead to a convergence?

Of course, in view of how far our actual position is from one in which we are acquainted with the totality of plausible moral conceptions and arguments, any answer that one gives to question (2) will be at least somewhat speculative. However, although the question cannot be definitely settled, we think that there are strong reasons to think that the answer to question (2) is 'No', beyond the simple plausibility considerations just mentioned. In particular, one thing that is quite suggestive in this context is the extensive and mathematically rigorous literature exploring the extent to which idealized Bayesian reasoners would converge in their beliefs over time when exposed to the same evidence.[17] Because we think that the parallel is illuminating in the present context, we would like to explore it at some length.

Like the reflective equilibrium theorist, the Bayesian takes to heart the lesson that, in deliberating about what to believe, we never 'start from scratch'; rather, we begin from a starting point that is not completely neutral among all possibilities. For the proponent of

[17] For a sophisticated overview of this literature, see Earman (1992) especially chapter 6.

Rawlsian reflective equilibrium, that starting point is a set of initial considered judgments; for the Bayesian, that starting point is some prior probability distribution. Given that orthodox Bayesians allow that even quite different prior probability distributions can be admissible starting points, the question can then be posed: to what extent would idealized Bayesian reasoners with different starting points converge over time, upon exposure to common evidence?

One thing that gives this question a certain urgency for many Bayesians is their claim that paradigmatic reasoning in the sciences is best understood in Bayesian terms. A natural and immediate challenge to this claim concerns whether Bayesians can account for the apparent objectivity of science, and the noticeable ability of various natural sciences to achieve consensus over time, given that the Bayesian will allow that individuals with different prior probability distributions might each be perfectly reasonable in holding quite different views on the basis of the same evidence. In this context, Bayesians sometimes take heart in a phenomenon known as the "swamping" of the priors. These convergence results (see, e.g., Doob 1971, Gaifman and Snir 1982) show that, for a relatively wide range of prior probability distributions, initial differences are swamped or washed out over time: as individuals are increasingly exposed to common evidence, their initial differences become increasingly insignificant, and they converge on a common view.

At first glance, the existence of such convergence results might seem highly encouraging for the reflective equilibrium theorist who thinks that it is important that there is a unique wide reflective equilibrium. For this seems to be a near perfect model for the kind of thing that she envisages: even significant differences among the initial considered judgments held by different individuals are eventually washed out as those individuals are increasingly exposed to “all feasible moral conceptions and all reasonable arguments for them”. However, we think that this is the wrong lesson to take away from the discussions of convergence in the Bayesian literature. Indeed, we think that the lessons of that literature should decrease, rather than increase, one’s confidence that there is a unique wide reflective equilibrium. First, we note a potentially crucial difference. For the orthodox Bayesian, there is a single, perfectly determinate norm that governs all belief revision: that of conditionalization. Whenever one acquires a new piece of evidence, one should update one’s prior opinions in accordance with Bayes’ theorem. In effect, given a prior probability distribution, there is no space for judgment about how one should respond to a newly-encountered piece of evidence; the uniquely rational response is already fixed by one’s prior commitments. But one might reasonably think that this is not how things are in the moral case. Rather, responding to a newly encountered moral consideration, argument or conception will require a certain amount of judgment; how one should respond is not simply given by one’s prior commitments.[18] And this already seems to introduce a level of potential slack in the reflective equilibrium picture that is not present in the Bayesian picture. In any case, proponents of the method have never proposed norms (let alone a single, master norm) for pursuing wide reflective equilibrium that has anything like the determinateness of Bayesian conditionalization.

[18] That this is so, at least for Rawls himself, is suggested by passages such as the following: “Moral philosophy is Socratic: we may want to change our present considered judgments once their regulative principles are brought to light. And we may want to do this even though these principles are a perfect fit. A knowledge of these principles may suggest further reflections that lead us to revise our judgments” (1971: 49).

But let us waive this potential difference. The crucial point is this: even if it were given that the application of the reflective equilibrium procedure leads to convergence results that are *as robust* as the kind of convergence produced by Bayesian conditionalization, this would not be enough, for it turns out that there are many admissible prior probability distributions that do *not* lead to convergence over time, no matter how much common evidence is provided to the inquirers. Here we should note a crucial similarity between the orthodox Bayesian and the proponent of Rawlsian reflective equilibrium: both seem to be extremely liberal in what can count as an admissible starting point. For the orthodox Bayesian, any prior probability distribution that satisfies certain purely formal constraints[19] is admissible; because of this, even radically different starting points are admissible. And it is this fact which guarantees that, in principle, two inquirers might fail to converge even in the hypothetical long run, despite the fact that they both begin from admissible prior probability distributions. Similarly, given Rawls' characterization of considered judgments, it is quite clear that different individuals might begin from radically different sets of judgments, all of which qualify as considered judgments. (Consider again the differences in the kinds of stable judgments that people make about the organ harvesting case, even when they are not upset or frightened, etc.).

In fact, there is an obvious respect in which the orthodox Bayesian is significantly *less* permissive in what he will allow as an admissible starting point than the proponent of Rawlsian reflective equilibrium. For the orthodox Bayesian will require that any admissible starting point is a *probabilistically coherent* set of credences—thus, any

[19] In particular, it is both necessary and sufficient that a prior probability distribution satisfies the axioms of the probability calculus.

starting point will contain no internal conflicts. By contrast, the Rawlsian reflective equilibrium theorist clearly will allow, among admissible starting points, sets of considered judgments that contain internal conflicts; indeed, explications of the method typically presuppose that any actual set of initial considered judgments will contain at least some such conflicts, for this is one of the primary factors that propels the process of revision forward.

On balance, and mindful of the limits of this kind of argument, we think that the investigation of convergence in the Bayesian literature suggests that

> There is not a unique wide reflective equilibrium across different individuals.

Put otherwise:

> Different individuals might impeccably employ the method of (Rawlsian) reflective equilibrium and end up with substantially different moral views, even if they were exposed to all feasible moral conceptions and all reasonable arguments for those conceptions.

Suppose that this is true. What would follow? Rawls himself seemed to think that the very *existence* of 'objective moral truths' presupposes that there is a unique wide reflective equilibrium, or at least, that any differences between moral views affirmed in wide reflective equilibrium would be relatively marginal. (1974:290, 301). Indeed, the fact that he repeatedly and quite self-consciously eschewed any talk of truth in the moral domain seems to have been at least in part due to this view, combined with increasing skepticism about whether diverse individuals competently pursuing reflective equilibrium would ultimately converge in their substantive moral views.[20] If this is correct, then it is

[20] The fact that diverse individuals cannot be expected to converge on the same substantive moral views in wide reflective equilibrium, and the consequences of this, is one of the driving themes of Rawls (1993).

obvious that the method of reflective equilibrium will not deliver moral knowledge, for moral knowledge requires moral truth. Interestingly, others, including some prominent moral realists, have similarly suggested that it is a necessary condition for the truth of moral realism that rational inquirers would converge on a common moral view (See, e.g., Smith 1994, 2000). If this is correct, and if the method of reflective equilibrium is in fact the correct methodology for investigating the moral domain, then the non-existence of a unique wide reflective equilibrium would entail the falsity of moral realism.

We think that Rawls was right to be skeptical about the existence of a unique wide reflective equilibrium but wrong to assume that moral realism (or "objective moral truths") requires this. As we have seen, David Lewis held that different philosophers, each of whom is pursuing equilibrium among her opinions in a rationally impeccably manner, might ultimately settle on different equilibria, but that this is no reason to doubt that there is an objective matter of fact that divides them. We believe that what Lewis thought was true of philosophy in general holds also for the moral domain: even if different individuals who have impeccably applied our best methods of moral inquiry arrive at incompatible views in wide reflective equilibrium, that does not entail that moral realism is false. Indeed, we find the suggestion that such an entailment holds somewhat puzzling. Of course, if one thought that the impeccable application of our best methods for investigating a given domain is guaranteed to deliver the truth about the domain (at least in the long run), then there would be a good reason to think that different inquirers would ultimately converge on a single view if they impeccably applied those methods. But as noted in Section 1, no realist should accept the assumption that the impeccable application of our best methods is guaranteed to lead to truth in the long run. Indeed, it has often been taken as definitive of a realist stance towards some domain that one thinks that a non-epistemic notion of truth is applicable to statements of that domain; accounts of truth that tie the notion closely to the deliverances of our epistemic procedures, even idealized versions of our epistemic procedures, are treated as paradigms of *anti*-realism.[21] And once it is admitted that even idealized inquiry in some domain might leave us short of the truth (as the realist about that domain supposes is possible), it is unclear what further reason there is to suppose that rational inquirers who begin with diverse commitments are guaranteed to converge on a single view. In the absence of a compelling argument for the claim that moral realism presupposes a unique wide reflective equilibrium, we should reject the alleged entailment.[22]

Still, even if the fate of moral realism does not hang in the balance, one might very well think that it is an objectionable feature of the method of reflective equilibrium if it allows for the lack of convergence, and perhaps even radical divergence, envisaged here. According to this line of thought, a good method for investigating a given domain should lead rational inquirers who impeccably follow that method to converge in their views over time. This certainly seems true of the methods employed in those domains where we are most confident that genuine knowledge is acquired as a result of systematic

[21] Two famous examples of the latter: C.S. Peirce's (1940) view that truth is the opinion on which scientists would converge in the hypothetical limit of scientific inquiry and Hilary Putnam's (1981) "internal realism", according to which truth is identical with rational acceptability in ideal epistemic conditions.

[22] Rawls himself does not offer an argument, but see Smith (1994, esp.ch.5, and 2000: 34-36). For criticism of Smith on this issue, see Enoch (2007); Smith (2007) is a reply to Enoch. Although we cannot pursue this issue any further here, we hope to return to it in future work.

inquiry, e.g., mathematics and certain empirical sciences. Of course, even if the method of reflective equilibrium is deficient in this respect compared to procedures that are available in certain domains, it might still be the best procedure that we have for philosophical or moral inquiry. Alternatively, a proponent of the method might respond to worries about a lack of convergence by offering a less liberal characterization of what constitutes an admissible starting point. This last possibility will loom large in the final two sections of the paper.

4. Scanlon

Scanlon's "Rawls on Justification" (2002) offers not only a sophisticated interpretation of Rawls' views on the method of reflective equilibrium, but also a formidable defense of that method. His defense occurs largely in the course of discussing three objections: that the method begs the question against moral skepticism (pp.145-147), that it is overly conservative (pp.150-151), and that it is relativistic in an objectionable way (pp.151-153). Here we will not attempt to do full justice to Scanlon's discussion, but rather selectively mine it with an eye towards those issues that emerged as pressing in our examination of Goodman and Rawls. In particular, we will consider Scanlon's responses to the concerns that (i) considerations of coherence cannot bear the kind of weight that the reflective equilibrium theorist seems to place on them, and (ii) worries that the method is relativistic in an objectionable sense, inasmuch as there is no reason to think that rational inquirers who employ the method will converge in their views.

Consider first a thoroughgoing skeptic about morality, who stands to non-skeptical first-order moral thought as the atheist stands to religious belief. If justification in the moral domain consists simply in pursuing and achieving an equilibrium among our considered judgments about morality, then it might seem that the non-skeptic is in a position to make suspiciously quick work of any challenge that the skeptic might offer. In justifying her non-skeptical judgments about particular cases, the non-skeptic will simply appeal to certain general principles that she also accepts. When it comes time to justify her commitments to those general principles, she will cite the fact that they account for and explain her considered judgments about cases. Offhand, this looks too easy. (Compare the sense that showing that certain particular inductive inferences are in accordance with our actual inductive practice seems to be a rather meager response to the inductive skeptic.)

Scanlon interprets this familiar worry in terms of a comparison with astrology:

> Suppose, for example, that we were to undertake to render into coherent form the judgments about astrology in which people felt most confidence, revising many of these judgments in the process. This would not allay reasonable doubts about whether astrology is something we should take at all seriously. The result would not be a set of justified astrological judgments but only, at best, a set of claims that was internally consistent. Similarly, it may be said, merely subjecting our considered judgments about morality to scrutiny and possible revision through the method of reflective equilibrium does not provide an adequate response to doubts about morality (pp.145-146).

In response, Scanlon notes a significant difference between astrology and morality: astrology, but not morality, is committed to causal claims about physics and psychology that are clearly false. Thus, even achieving perfect coherence among our astrological judgments would not undermine the strong reasons that we have for skepticism about astrology. In contrast, Scanlon, following Rawls, holds that morality makes no claims that are even potentially contradicted by physics, psychology, or any other empirical science. Indeed, according to Scanlon, morality has no “external commitments” at all. By this, he means that the reasonableness of our taking moral judgments seriously does not depend on any claims that extend beyond morality itself (p.146).

One might worry that Scanlon artificially weakens the challenge by taking astrology as his foil. Consider a potentially more difficult comparison: theology, understood as "transcendent metaphysics". In the twentieth century, various philosophers who addressed the status of religious claims treated such claims as instances of what they called *transcendent metaphysics*: claims that have absolutely no observable consequences or upshot for the empirical world.[23] On this understanding of claims like "God exists", whether this claim is true or false makes absolutely no difference to anything that we observe, or anything that could even in principle be detected by the empirical sciences; for this reason, positivists like Ayer declared that such claims were neither true nor false, but rather cognitively meaningless. We think that this understanding of religion is a peculiar one, inasmuch as we suspect that there have been relatively few religious believers whose beliefs are plausibly interpreted in this way.[24] Nevertheless, let us consider the case of a religious believer whose theological commitments are best interpreted as pieces of transcendent metaphysics; *pace* Ayer and other logical positivists, we will also assume that claims like 'God exists' in the believer's mouth are meaningful, and hence capable of being either true or false. Suppose that many of the believer's commitments strike us as utterly fantastic and bizarre. Nevertheless, he manages to bring

[23] Ayer (1936, ch.VI) is a *locus classicus* of the genre.

[24] After all, even Enlightenment deists, notable among believers for the extent to which they held that God does *not* actively intervene in the empirical world, typically believed that God was responsible for creating the world in the first place. Even on this view then, it is certainly not the case that human experience would be no different if not for God's active intervention.

them into perfect equilibrium; indeed, we can imagine that his theological beliefs possess a level of internal coherence that far surpasses that which obtains among our beliefs in many domains where we nevertheless take ourselves to have genuine knowledge. Still, our doubts about his theological beliefs might persist. Of course, if we expressed doubt about any particular belief, he would be able to cite other beliefs that he holds, which stand in certain logical and quasi-logical relations of support to it. Nor could we appeal to any facts from other domains that contradict (or even stand in tension with) these theological commitments. For because the theological system has the status of transcendent metaphysics, it does not make any claim that could even in principle be contradicted by some empirical observation or scientific theory. (In this respect, it is quite different from astrology.) Indeed, because it is a piece of transcendent metaphysics, the theology has no "external commitments" at all: like morality as understood by Scanlon and Rawls, it is simply its own self-contained subject matter.

We submit that it does not follow that, in these circumstances, the believer's considered judgments about theology are justified. If skeptical concerns were raised about those commitments--how, after all, does he know that the whole thing is not simply a coherent fantasy?--and he responded by demonstrating that particular commitments cohere well with others, his response would be inadequate. More generally, in order for one's considered judgments about some domain to be justified, it is not enough that those judgments (i) cohere well with one another and (ii) are not contradicted by judgments from outside that domain. But if this is right, then it follows that it is not enough for our moral judgments to be justified, that they cohere well with one another and are not contradicted by well-confirmed views from outside of morality.

How might Scanlon respond to the theology comparison? We suspect that he would reply along the following lines. Although theology-as-transcendent-metaphysics resembles morality in that neither has any *empirical* presuppositions, of a kind that might be contradicted by the sciences, it does (obviously enough) have controversial *metaphysical* presuppositions. But Scanlon, following Rawls, holds that morality has no controversial metaphysical presuppositions (p.146). And this is because, on Rawls' view, "the presuppositions that need to be redeemed to defend morality are practical rather than theoretical" (146). Here, the constructivist aspects Scanlon's conception of morality come to the fore, and the categorization of that conception as realist becomes at least somewhat tenuous. Indeed, it is perhaps significant that Scanlon, like Rawls, studiously avoids any talk of *truth* in his defense of the method of reflective equilibrium.[25] Since our concern is with the method of reflective equilibrium as a tool for the discovery of truth, we will not pursue the argument any further, beyond noting the following: even if Scanlon is correct in thinking that the method of reflective equilibrium is impervious to the charge that it "begs the question against skepticism" when it is applied to domains that have no substantive empirical or metaphysical presuppositions, it does not follow that the same objection misses the mark when the method is used to investigate domains

[25] The closest surrogate notion for Scanlon seems to be that of being *justified*, in an impersonal sense: in this sense, a moral principle or judgment is justified when "it is supported by good and sufficient reasons" (p.140). This contrasts with a weaker sense of justification, in which a *person* might be justified in believing a moral principle. Thus, that a person is justified in believing a moral principle does not entail that the principle itself is justified: "A person can be justified…in accepting a principle (for certain reasons) even though the principle itself is not justified because, say, there are other factors (of which we he could not be expected to be aware of) that undermine the justificatory force of the considerations he takes to be reasons for it" (140).

that do have substantive empirical or metaphysical presuppositions. (Of course, many will think that morality itself is one such domain.)

Let us turn to Scanlon's discussion of the possibility that there is no unique wide reflective equilibrium (pp.151-153). Scanlon is prepared to admit, at least for the sake of argument, that equally well-informed people might carry through the process equally conscientiously and yet arrive at different equilibria. Is the proponent of reflective equilibrium committed to the view that both individuals are justified in holding their differing views in the circumstances? According to Scanlon, she is not. He discusses the case from the first person perspective, as one of the individuals who has achieved equilibrium:

> Faced with the case of someone who reaches an equilibrium different from my own, I must ask why this divergence occurred. If it occurred because the person began with different considered judgments, I must ask whether I think, on further reflection, that the judgments that person accepted are correct and whether he or she was correct in rejecting ones that I accepted…If the divergence occurred because the person made different choices at later stages in the process…then I need to consider whether these decisions were reasonable…The reexamination provoked by a case of this kind may disrupt the equilibrium I had reached, but it need not do so (p.152-153).

Suppose that the divergence is due to the fact that the two parties set out from different starting points. In that case, Scanlon suggests that one should ask whether the considered judgments from which the other person begins are *correct*; presumably, the thought is that if one answers this question negatively, then that will provide a reason for downgrading the significance of the fact that she has reached a different equilibrium. However, such a maneuver seems problematic. For given the picture of the method typically offered by its proponents, it is extremely likely that, when one judges one's *own* starting point from the perspective of wide reflective equilibrium, one will conclude that a significant number of the considered judgments that one held then are false. Indeed, defenders of the method frequently emphasize the possibility and likelihood of dropping a significant number of one's initial considered judgments, in the context of attempting to parry the charge that the method is overly conservative and unduly privileges the considered judgments with which one begins. So the fact that the starting point of the Other seems to contain a significant amount of error, when judged by one's current lights, will not typically distinguish it from one's own starting point.[26]

In what sense might the Other's starting point be defective compared to one's own, given that it is quite likely that each contains a substantial amount of falsehood from the perspective of wide reflective equilibrium? The obvious answer is that some sets of initial considered judgments are more reasonable, or have greater rational credibility, than others. If one judges that the initial considered judgments of the Other were on the whole less reasonable than one's own initial judgments were, then that would seem to break the otherwise threatening symmetry, at least if one's current judgment to that effect is correct.[27]

In fact, this seems to be Scanlon's view. In a footnote attached to the paragraph from which the above passage is drawn, Scanlon addresses Richard Brandt's claim that, if the

[26] Of course, there is the following asymmetry: one's own starting point, when transformed by a conscientious application of the method of reflective equilibrium, has led to one's current view (which, trivially, one now takes to be correct), while the starting point of the Other failed to do so. But surely to appeal to the correctness of one's current view in order to bolster the comparative rational credentials of one's starting point really is viciously circular, given that one intends to appeal to the superior rationality of one's starting point to justify maintaining one's current view.

[27] On the importance of the correctness of such judgments in breaking otherwise threatening epistemic symmetries, see Kelly (2005, 2010, forthcoming).

method of reflective equilibrium is to lead to beliefs that are justified, then some of the beliefs with which it begins must be "initially credible—and not merely initially believed—for some reason other than their coherence" (Brandt 1979:20). In response, Scanlon says the following:

> Thus, Brandt…was correct that the justificatory force of an application of the method of reflective equilibrium depends on the credibility of its starting points…But that is not an objection to that method (p.167).

Here we should note two points that have emerged. First, if what Scanlon says in response to Brandt is correct, then some starting points can have credibility that others lack despite the fact that the judgments that constitute those starting points all qualify as considered judgments, in the honorific sense. That is, it is not enough that a judgments satisfies the condition for being a considered judgment that it possess rational credibility, for some considered judgments lack credibility.[28] Secondly, given that Scanlon presents himself as conceding Brandt's point (although denying that it amounts to an objection), it seems that the fact that some considered judgments possess "initial credibility" is not a matter of their standing in relations of coherence to other beliefs.

We believe that Scanlon is right to hold, with Brandt, that whether an application of the method of reflective equilibrium yields justified beliefs depends on the credibility of the starting points from which it begins. But we suspect that this is a greater concession

[28] On p.14, Scanlon comes close to endorsing the view that all considered judgments, as such, have initial credibility: "…the judgments that meet these conditions [i.e., the conditions for being a considered judgment] state those things that seem to us most clearly to be true about moral matters if anything is, and...unless there is some ground for doubting them it is reasonable to grant them initial credibility". But notice that even here, the hedge "unless there is some ground for doubting them" seems to suggest that fulfilling the conditions for being a considered judgments is *not* a sufficient condition for possessing initial credibility.

to critics of the method than Scanlon thinks. This is among the issues that we will explore in the final section of the paper.

5. Is Reflective Equilibrium Enough?

In Section 1, we noted that a common charge among detractors of the method of reflective equilibrium is that the method is extremely *weak*: that is, they claim that, even if one impeccably executes the method, one might very well arrive at views that lack various desirable features. A striking fact that has emerged from our survey of Goodman, Rawls, and Scanlon is the extent to which they seem to share this outlook. That is, it is striking how modest they are about what status can be claimed for deliverances of the method. For Goodman, even when the method is impeccably applied to arrive at true beliefs about the future, those beliefs will inevitably fall short of being knowledge. Rawls was agnostic (at his most optimistic) and skeptical (in his later writings) that diverse individuals who applied the method would converge on a unique wide reflective equilibrium, something that he regarded as a necessary condition for the very existence of "objective moral truths" (and therefore, presumably, for knowledge of such truths). Scanlon acknowledges that the "justificatory force" of an application of the method depends on the credibility of its starting points. Thus, Scanlon seems to agree that even if a person begins from all and only her considered judgments, and then successfully achieves wide reflective equilibrium, her views might nevertheless not be justified, if the considered judgments from which she sets out are sufficiently lacking in rational credibility.

Let us dwell on this last point. In Section 1, we argued that, although it is not an objectionable feature of a method that it could lead one to views that are *false*, it *is* a good objection to a method if it turns out that impeccably following that method could lead one to views that are *unreasonable*. It follows from this that if beginning from all and only one's considered judgments, and from there achieving wide reflective equilibrium without making any "downstream" mistakes, is sufficient for impeccably executing the method of reflective equilibrium, then the method is not correct. The problem is that something might very well qualify as a considered judgment, when that notion is understood in anything like the way it is understood in the broadly Rawlsian tradition, and yet be utterly lacking in rational credibility. For example, given a standard Rawlsian characterization, there is in principle nothing that precludes the following from qualifying as a considered judgment for someone:

> One is morally required to occasionally kill randomly.

For there is nothing *incoherent* about the possibility that someone could confidently and stably subscribe to this judgment, even if he or she is aware of all of the non-moral facts, does not stand to gain or lose depending on whether it is true or false, and so on. Not only do such conditions fail to preclude the possibility of someone's having this among his or her considered judgments, but it seems that such a proposition might score just as high along the relevant dimensions as the following proposition does for the average person who believes it:

> One is not morally required to occasionally kill randomly.

The weakness of the conditions typically imposed on “considered judgments” is sometimes obscured by the choices of examples that are given when the method of reflective equilibrium is illustrated by its proponents. The theorist illustrating the method typically proceeds from the first person perspective, and speaks of (e.g.) “our” considered judgments; she thus selects one of her own considered judgments that she expects her readers to share. This is of course natural enough—certainly, it would be strange for a proponent to illustrate the method by proceeding from a judgment that she does not hold, or which she does not expect her audience to share. But nevertheless, it is a dangerous procedure. For proceeding in this way runs the risk that what we are responding to, in agreeing that a certain judgment is part of an appropriate starting for conducting inquiry, is *not* its status as a considered judgment, but rather our perception that it has some more significant positive epistemic status: for example, that it has a high degree of rational credibility, or even that it is among the things that we know to be true. How might one discriminate between these two possibilities? In order to test the claim that it is the fact that the judgment in question is a considered judgment which is doing the work in this context, it is important to consider cases from the third person perspective, in which the starting points of the person pursuing reflective equilibrium are (i) his considered judgments but (ii) *perverse* considered judgments, at least when judged by one’s own lights. (That is, judgments which, when judged by one’s own lights, are clear cases of non-knowledge, or propositions utterly lacking in rational credibility.) We think that when one performs this experiment, the idea that the normatively appropriate starting point for a person consists of all and only her considered judgments increasingly loses its appeal.

Suppose that proponents of reflective equilibrium simply *dropped* the apparatus of considered judgments, understood in anything like the way that notion is understood in the broadly Rawlsian tradition, and appealed directly to judgments that have some *substantive* positive epistemic status. For example, suppose that instead of endorsing (1) a proponent endorsed (2):

(1) For any individual, the appropriate starting point from which to pursue wide reflective equilibrium is the class of judgments consisting of all and only her considered judgments.

(2) For any individual, the appropriate starting point from which to pursue wide reflective equilibrium is the class of all and only those judgments that she is justified in holding at that time.

Of course, someone might claim that (1) and (2) come to the same thing, on the grounds that a person is justified in holding a judgment prior to beginning the pursuit of reflective equilibrium just in case that judgment is among her considered judgments. Again, we think that that is a mistake: it does not follow that a person is justified in believing that *we are under a standing moral obligation to occasionally kill randomly*, even if that belief is among her considered judgments. In any case, it is clear that theorists like Scanlon do not take all considered judgments to have the same level of initial credibility. We will assume then, that (1) and (2) are distinct alternatives.

We have already presented one reason for interpreting the method in terms of (2) rather than (1), viz. that (1) seems overly inclusive. It is worth noting that (1) also seems to suffer from the opposite problem: that of being overly exclusive. That is, if one restricts the starting point to *only* considered judgments, then certain judgments might be excluded from playing a role in subsequent deliberation, in virtue of failing to qualify as 'considered', despite the fact that, intuitively, the judgments in question *ought* to play a role in one's deliberations about which theory to accept.

Rawls, for example, suggests that one set aside judgments that are heavily bound up with one's own interests. But consider the proposition that

> A person of color should not receive lesser consideration in virtue of being a person of color.

Notice that, for a person of color, this judgment is heavily bound up with his or her own interests.[29] Offhand, it seems like this judgment might fail to qualify as a considered judgment for a person of color for that reason, and thus should be excluded from her subsequent deliberations. This seems like the wrong result, however. On the contrary, we think that it would be perfectly reasonable for a person of color to give a great deal of weight to this proposition in working towards a reflective equilibrium.[30] We think that the reason for this is the following: despite the fact that it is very much in the self-interest of a person of color that this proposition is true as opposed to false, she will still typically have a high degree of justification for her belief that it is true; because of this, not only is it rationally permissible for her to take this proposition into account in her deliberations, but it would be a mistake for her to set it aside.

Notice that both the "overly inclusive" and "overly exclusive" problems for (1) issue from the same source. Considered judgments are ones that are held in conditions that are

[29] Of course, we do not mean to suggest that this proposition is not bound up with the self-interests of others as well.

[30] We do not mean to suggest that Rawls himself would disagree with this verdict, only that, given the conditions that he offers, its status as a considered judgment is at the very least problematic. Moreover, even if there are resources within the Rawlsian account for blocking this particular example, the underlying problem (which we spell out in the next paragraph) would persist.

hospitable or *conducive* to judging well. However, even if one is in conditions that are hospitable or conducive to good judgment, there is no guarantee that the judgment at which one arrives will be reasonable as opposed to unreasonable. Not all unreasonableness is due to the operation of the kind of general corrupting factors (e.g., being personally invested in how a given question is answered) that the relevant conditions exclude. Conversely, even when one is in conditions that are in some respects *inhospitable* to good judgment, there is no guarantee that the judgment at which one arrives will be less than perfectly reasonable. Here as elsewhere, there is a substantive gap between the quality of the conditions under which one performs and the quality of one's performances: both good performances in adverse conditions and bad performances in favorable conditions are eminently possible, even if less likely than other combinations. This element of slack causes difficulties for the view that all and only one's considered judgments constitute the appropriate starting point for inquiry. For in deciding which theory to accept, it will seem wrong to give weight to unreasonable considered judgments, and (perhaps even more clearly) wrong to give no weight to perfectly reasonable judgments that do not qualify as 'considered'. Moreover, it would be a mistake to think that the line of criticism developed here depends on some particular or idiosyncratic conception of what it is to be a considered judgment, as though the difficulties could be avoided by (e.g.) tweaking the conditions offered by Rawls or Scanlon. Although of course particular counterexamples can be blocked by amending the conditions in various ways, the underlying problem is a more general one, and will arise for any account that explicates "considered judgment" in terms of conditions that are only *generally* hospitable or conducive to good judgment. By contrast, an account that appeals directly to the normative status of the judgments themselves, such as (2), does not suffer from the relevant problems.

Thus, we think that incorporating something along the lines of (2) into one's account of the method yields a more defensible account than incorporating (1). Nevertheless, the latter understanding of the method seems much more popular among its proponents. This raises the question of why theorists have wanted to understand the method along those lines. We can think of several motivations for preferring (1) over (2). These divide into two main classes: *epistemological* motivations and *metaphysical* motivations. Although we think that neither kind of motivation is compelling, we want to consider each carefully.

Consider first *epistemological* motivations for preferring a characterization of the proper starting point in non-normative terms. On its standard, deliberative interpretation, the method of reflective equilibrium purports to be a method of belief revision that is suitable for guiding an inquirer: the method purports to be a (non-algorithmic) decision procedure that one can apply from the inside in order to figure out what to believe. It is a notable feature of the conditions that Rawls lays down for being a considered judgment that it is typically quite *easy to tell* whether a given judgment satisfies these conditions. For example, one is typically in a good position to tell whether one is frightened or upset when thinking about a question, whether one is relatively confident as opposed to uncertain whether a given judgment is true, and so on. Thus, it seems that one would generally be quite reliable in determining whether a given judgment qualifies as *considered* or not. Given that the method of reflective equilibrium requires one to be able to reliably identify considered judgments as such, it seems like the task will be an eminently manageable one. On the other hand, one might worry about one's ability to successfully follow a method that requires one to identify those judgments that have some more objective positive epistemic status. How after all, is one supposed to assess the initial, rational credibility of some proposition, or how well justified one's belief would be if one came to believe that proposition?

The worry becomes more salient in a context of interpersonal disagreement, in which others are inclined to take issue with one's assessment. For example, perhaps at the outset of inquiry you are inclined to give significant weight to your considered judgment that

> Even if a doctor could save two lives by forcibly harvesting the organs of some unwilling bystander, he is not morally required to do so.

However, the act utilitarian thinks you should give little or no weight to this judgment: as far as he is concerned, it is for all you know simply a blind and baseless prejudice, and so should be set aside when it comes to deciding which theory to accept. But notice that if *being a Rawlsian considered judgment* is sufficient for being properly taken into account, then it seems as though you are in a strong position to establish the propriety of your treating this proposition as a relevant consideration, perhaps even to the satisfaction of a reasonable act utilitarian. After all, who is the act utilitarian to deny that this is among your considered judgments? You introspect carefully and find yourself quite confident, neither frightened nor upset, and committed to the judgment in a way that is stable rather than fleeting. Moreover, you think that, in the unlikely event that such a scenario were to actually arise, you would be no more or less likely to find yourself in the role of the innocent bystander than in the role of dying patient; you are thus not invested in the question in a self-interested way. Recognizing that you satisfy the other conditions as well, you conclude that the judgment in question really is among your considered judgments. Surely the act utilitarian is not in a good position to take issue with *that* assessment. On the other hand, if you were to claim that the reason why you intend to give significant weight to this judgment in deciding which moral theory to accept is because it has a high degree of rational credibility, or because you are currently justified in holding it, then the act utilitarian will immediately object and accuse you of begging the question against his theory. Thus, one might be led to understand the method in terms of (1) rather than (2) because doing so makes it easier to *follow* the method, and to establish to the satisfaction of others that one is correctly following it.

While we appreciate the pull of this line of thought, we think that it should be resisted. Notice first that, although individuals will generally be quite good at identifying their considered judgments as such, they will also be eminently *fallible* executors of the task. (And this will be so for any explication of 'considered judgment' that has yet been proposed.) Indeed, on any plausible explication, errors will be possible in both directions: one might mistakenly take something to be a considered judgment that fails to satisfy at least one of the conditions, and one might also fail to recognize something that is in fact a considered judgment as such. This fallibility with respect to questions about what our considered judgments are follows from our fallibility with respect to questions about whether the relevant conditions obtain in particular cases.

Consider next a version of the method which characterizes the starting point in explicitly normative terms, e.g., as those propositions that have sufficiently high initial credibility, or those propositions that one justifiably believes as the search for reflective equilibrium begins. In order to correctly apply this version, one will have to possess an at least reasonably competent grip on which propositions are initially credible; if we were hopelessly unreliable or at sea with respect to such assessments, then, given that the method is supposed to be something that we can apply, this would not be a viable construal. But in fact, setting aside a radical and as yet unmotivated skepticism, we are not hopelessly unreliable with respect to such questions. One can recognize that some propositions about what one is morally required to do have greater initial credibility than others. Granted, given the relative ease with which considered judgments can be identified as such, we would expect people to make more mistakes about which propositions are rationally credible than about whether something is among their considered judgments. But given that the difference is a matter of degree as opposed to kind, it is unclear why that should motivate interpreting the method in terms of (1) rather than (2), in light of the problems that beset the method when it is interpreted in that way.

Consider how the difference between (1) and (2) plays out with respect to concerns about self-interest. Above, we suggested that, even if it is very much in one's self-interest that p is true rather than false (where p is some moral claim), it does not follow that one should set aside one's judgment that p in evaluating rival moral theories. This is because, even if one is in less than ideal circumstances for making the judgment, and one appreciates this fact, one's judgment might nevertheless be justified. Of course, the fact that one should not automatically set aside any moral judgment that strongly aligns with one's self-interest will make it significantly more difficult in practice to correctly manage one's biases: from the inside, a case in which one unjustifiably holds a belief because it aligns with one's self-interest might feel very much like a case in which one is genuinely justified in holding a belief that happens to align with one's self interest. A norm such as "One should set aside any moral judgment that strongly aligns with one's self-interest" would have us treat these difficult to distinguish cases alike; it is in that respect relatively easy to apply. (Although again, we should expect honest failures of compliance to occur even with respect to such easy to apply norms.) But that is a poor reason for adopting this as a norm of inquiry, as opposed to a norm of inquiry that would have us distinguish, among judgments that align with our self-interest, between those that are justified and those that are not.[31]

Although we think that this kind of epistemological motivation plays a role in the tendency to characterize the correct starting point in non-normative terms, we suspect that another kind of motivation is perhaps even more important. One of the things that many philosophers have found quite appealing about the method of reflective equilibrium is that it seems to provide a relatively down-to-earth epistemology for domains where the traditional alternatives have often seemed uncomfortably exotic. As we have noted, the method has generally been most popular as an account of the methodology for domains in which our knowledge is not in any obvious way underwritten by sense perception. If it is assumed that moral knowledge, or philosophical knowledge, or knowledge of logic, is not underwritten by sense perception, then one might be tempted to postulate (e.g.) a *sui generis* faculty of rational intuition, or some other surprising mechanism, in order to play the underwriting role that sense perception plays with respect to straightforwardly

[31] Williamson (2000) emphasizes that it is a perennial temptation in philosophy to *interiorize* the application conditions for norms, methods, and procedures. As he trenchantly argues, however, the motivations for doing so tend to collapse when subjected to close scrutiny. In his later (2007 ch.7), he provides a compelling critique of “evidence neutrality”, or the idea that whether a proposition constitutes evidence for an inquiry is in principle uncontentiously decidable among those engaged in the inquiry.

empirical knowledge. And of course, temptations of this general sort have often been indulged in the history of philosophy.

In contrast, the account of justification provided by the reflective equilibrium conception seems, in Goodman's phrase, "refreshingly non-cosmic" (p.60). Thus, when Goodman first introduces the method in the context of discussing the justification of deduction, he contrasts it with appeals to self-evidence and accounts on which the relevant logical principles are "grounded in the very nature of the human mind". By contrast, on the account of justification that he will offer, the truth lies "much nearer the surface" (p.61). Traditional attempts to justify induction suffer from a similarly fantastic character:

> The typical writer begins by insisting that some way of justifying predictions must be found; proceeds to argue that for this purpose we need some resounding universal law of the Uniformity of Nature, and then inquires how this universal principle itself can be justified. At this point, if he is tired, he concludes that the principle must be accepted as an indispensable assumption; or if he is energetic and ingenious, he goes on to devise some subtle justification for it. Such an invention, however, seldom satisfies anyone else…(pp.61-62).

But again in marked contrast, the account of justification that he will offer has nothing fantastic about it.

The same theme looms large in ethics. Thus, Norman Daniels (1996) one of the most ardent defenders of the method among moral philosophers, repeatedly claims that it is a great advantage of the method that it allows us to dispense with any need to postulate faculties of moral intuition, or similar exotica. One can certainly appreciate the appeal of this idea: after all, there is nothing remotely strange or exotic about the practice of seeking to make one's existing beliefs more coherent; indeed, on anyone's account, this is presumably a practice in which people frequently engage. Moreover, there is nothing far-fetched about the idea that we could succeed in making our moral beliefs coherent, or at least, significantly more coherent than they currently are. If this is what justification in the moral sphere amounts to, then such justification seems at least in principle within our grasp, and not something that requires the operation or even existence of dubious faculties, or some of way of making sense of the idea that certain privileged propositions have a higher ‘intrinsic credibility’ than others, independently of considerations of coherence.

Indeed, as standardly presented the method of reflective equilibrium is sufficiently unpretentious that it might seem to constitute less of an *alternative* to traditional accounts than a way of *evading* a certain kind of traditional epistemological puzzle: that of showing how justification is possible in a domain where it seems that sense perception is not playing its customary and familiar role. For example, if, as Lewis suggests, the method is the correct account of philosophy inquiry, then it seems as though more substantial attempts to account for philosophical knowledge along the lines of (e.g.) Bealer (1998) or Williamson (2007) are at best superfluous and at worst wrong-headed.

Notice, however, that if we interpret the method in such a way that correctly applying it requires that one *already* has some justified beliefs about the domain (or at least, that some starting points are more rational than others), then this ostensible virtue of the method seems to vanish. For in that case, there must be some *other* story, not itself provided by the reflective equilibrium picture, about why one is initially justified in believing certain things rather than others about the domain, or why certain privileged propositions have relatively high initial credibility. (Compare: on the assumption that we have at least some genuine knowledge of the future, and that Goodman was right to think that justification as he understands it is insufficient to underwrite such knowledge, then there must be some other story, waiting to be told, about how such knowledge is possible.) Consistent with the idea that we should strive to achieve reflective equilibrium among our beliefs, the need for a certain kind of traditional epistemological theorizing (with all of its attendant pressures towards postulating non-obvious normative mechanisms, and so on) seems to have re-emerged.[32]

We have argued that, in order to arrive at a defensible account, the proponent of the method should opt for a normative characterization of the starting point. Of course, once that move is made, one might very well wonder whether the picture of inquiry that emerges still deserves the name "the method of reflective equilibrium". For in that case, it is natural to think that the most interesting part of the story concerns not the pursuit of equilibrium itself, but rather what makes it the case that certain starting points are more reasonable than others, and how we manage to recognize or grasp such facts. In that sense at least, it seems that reflective equilibrium is not enough.[33]

[32] Indeed, it is quite possible that Rawls himself might have conditionally agreed with such an assessment of the situation. One of the tasks of "The Independence of Moral Theory" is to show that the pursuit of moral theory does not require one to engage in substantive epistemology. However, according to Rawls, what secures this independence from epistemology is precisely the fact that the primary aim of the reflective equilibrium procedure is not the discovery of objective truths (For this point, see pp.289-290). Given Rawls' non-veritistic conception of the goals of moral theory, his view that substantive epistemology is avoidable is perhaps unassailable. It is at any rate much less clear that this is true for those who embrace a more realist conception of the moral domain.

[33] This paper grew out of a graduate seminar that we co-taught at Princeton University in the Spring of 2010. We are grateful to the participants in that seminar for their feedback, and also to Tristram McPherson and Errol Lord for written comments.

References

Ayer, A.J. (1936). *Language, Truth, and Logic*. (New York: Dover).

Bealer, George (1998). "Intuition and the Autonomy of Philosophy". In Michael R. DePaul and William Ramsey (eds.) *Rethinking Intuition.* (Lanham, MD: Rowman and Littlfield): 201-240.

Blanshard, Brand (1939). *The Nature of Thought* (London: Allen and Unwin).

BonJour, Laurence (1985). *The Structure of Empirical Knowledge*. (Cambridge, MA: Harvard University Press).

BonJour, Laurence, and Sosa, Ernest (2003). Epistemic Justification. (Malden, MA: Blackwell Publishers).

Brandt, Richard (1979). *A Theory of the Good and the Right*. (Oxford: Oxford University Press).

Brandt, Richard (1990). "The Science of Man and Wide Reflective Equilibrium". *Ethics* 100 (2):259-278.

Cohen, Stewart (1984). "Justification and Truth". *Philosophical Studies* 46: 279-96.

Copp, David (1985). "Considered Judgments and Justification: Conservatism in Moral Theory". In D. Copp and M. Zimmerman (eds.) *Morality, Reason, and* Truth (Totowa, NJ: Rowman and Allenheld): 141-169.

Cummins, Robert (1998). "Reflection on Reflective Equilibrium". In Michael R. DePaul and William Ramsey (eds.) *Rethinking Intuition.* (Lanham, MD: Rowman and Littlefield): 113-127.

Daniels, Norman (1996). *Justice and Justification: Reflective Equilibrium in Theory and Practice*. (Cambridge: Cambridge University Press).

Daniels, Norman (2003). "Reflective Equilibrium". In Edward Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. URL: http://plato.stanford.edu/entries/reflective-equilibrium/

DePaul Michael (1998). "Why Bother With Reflective Equilibrium?" In Michael R. DePaul and William Ramsey (eds.) *Rethinking Intuition.* (Lanham, MD: Rowman and Littlfield): 293-309.

DePaul, Michael (2006). "Intuitions in Moral Inquiry", in David Copp (ed.) The Oxford Handbook of Ethical Theory. (Oxford: Oxford University Press): 595-623.

Doob, J.L. (1971). “What is a Martingale?”, *American Mathematical Monthly* 78: 451-462.

Earman, John (1992). *Bayes or Bust?* (The MIT Press: Cambridge, MA).

Enoch, David (2007). “Rationality, Coherence, Convergence: A Critical Comment on Michael Smith’s Ethics and the A Priori. *Philosophical Books* 48: 99-108.

Feldman, Richard (2003). *Epistemology* (Prentice Hall: Upper Saddle River, NJ).

Gaifman, H., and Snir, M. (1982). “Probabilities over Rich Languages”, *Journal of Symbolic Logic* 47: 495-548.

Gettier, Edmund (1963). “Is Justified True Belief Knowledge?”. *Analysis*.

Goodman, Nelson (1953). “The New Riddle of Induction”. In his *Fact, Fiction, and Forecast*. (Harvard: Harvard University Press).

Hare, R.M. (1973). “Rawls’ Theory of Justice”. *Philosophical Quarterly*, 23: 144-155; 241-251.

Harman, Gilbert (2004). “Three Trends in Moral and Political Philosophy” in *The Journal of Value Inquiry* 37: 415-425.

Hempel, Carl (1934-35a). “On the Logical Positivists’ Theory of Truth”. *Analysis* 2: 49-59.

Hempel, Carl (1934-35b). “Some Remarks on ‘Facts’ and Propositions”. *Analysis* 2: 93-96.

Kelly, Thomas (2005). “The Epistemic Significance of Disagreement”. In Tamar Szabo Gendler and John Hawthorne (eds.) *Oxford Studies in Epistemology*, vol.1 (Oxford: Oxford University Press): 167-196.

Kelly, Thomas (2010). “Peer Disagreement and Higher Order Evidence”. In Richard Feldman and Ted Warfield (eds.) *Disagreement* (Oxford: Oxford University Press).

Kelly, Thomas (forthcoming a). “Disagreement and the Burdens of Judgment”. In David Christensen and Jennifer Lackey (eds.) *The Epistemology of Disagreement: New Essays*. (Oxford University Press).

Kelly, Thomas (forthcoming b). “Following the Argument Where It Leads”. *Philosophical Studies.*

Lewis, David (1983). *Philosophical Papers, Volume I* (Oxford: Oxford University Press).

Lyons, David (1975). “Nature and Soundness of the Contract and Coherence Arguments”. In Norman Daniels (ed.) *Reading Rawls* (New York: Basic Books):141-167.

McMahan, Jeff (2000). “Moral Intuition”. In Hugh LaFollete (ed.) *The Blackwell Guide to Ethical Theory* (Malden, MA: Blackwell): 92-110.

Parfit, Derek (1984). *Reasons and Persons*. (Oxford: Oxford University Press).

Peirce, C.S. (1940). *The Philosophy of Peirce: Selected Writings*.

Plantinga, Alvin (1993). *Warrant: the Current Debate.* (Oxford: Oxford University Press).

Pollock, John and Cruz, Joseph (1999). *Contemporary Theories of Knowledge* (Boston, MA: Rowman and Littlefield).

Putnam, Hilary (1981). *Reason, Truth, and History*. (Cambridge: Cambridge University Press).

Rawls, John (1951). “Outline of a Decision Procedure for Ethics”. *Philosophical Review,* 60: 2:177-97. Reprinted in Rawls (1999): 1-19.

Rawls, John (1971). *A Theory of Justice*, 2nd Edition 1999, Cambridge, MA: Harvard University Press.

Rawls, John (1974). ‘The Independence of Moral Theory’, *Proceedings and Addresses of the American Philosophical Association* 47:5-22. Reprinted in Rawls (1999): 286-302. Page references are to the reprinted version.

Rawls, John (1980). “Kantian Constructivism in Moral Theory”. *Journal of Philosophy* 77: 515-572. Reprinted in Rawls (1999): 303-358. Page references are to the reprinted version.

Rawls, John (1993). *Political Liberalism*. (New York: Columbia University Press).

Rawls, John (1999). *Collected Papers*, Sam Freeman, (ed.), Cambridge MA: Harvard University Press.

Rawls, John (2001). *Justice as Fairness: A Restatement*, Cambridge MA: Harvard University Press.

Scanlon, T.M. (2002). ‘Rawls on Justification’, in Samuel Freeman (ed.) *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press): 139-167

Reichenbach, Hans (1938). *Experience and Prediction*. (Chicago, IL: University of Chicago Press).

Singer, Peter (1974). “Sidgwick and Reflective Equilibrium”. In the *Monist* 58: 490-517.

Smith, Michael (1994). *The Moral Problem*. (Malden, MA: Blackwell).

Smith, Michael (2000). “Moral Realism”. In Hugh LaFollete (ed.) *The Blackwell Guide to Ethical Theory* (Malden, MA: Blackwell): 15-37.

Smith, Michael (2007). "In Defence of *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics*: Reply to Enoch, Heironymi, and Tannenbaum". In *Philosophical Books*, 48: 136-149.

Stitch, Stephen (1990). *The Fragmentation of Reason*. (Cambridge, MA: MIT Press).

Timmons, Mark. (1987). “Foundationalism and the Structure of Ethical Justification”. *Ethics* 97 (3): 595-609.

Williamson, Timothy (2000). *Knowledge and Its Limits*. (Oxford: Oxford University Press).

Williamson, Timothy (2007). *The Philosophy of Philosophy*. (Malden, MA: Blackwell).