REFLECTIVE EQUILIBRIUM AND JUSTIFICATION
Daniel Little
Colgate University
Creighton Club
1981

Take it as given that the purpose of justification in moral philosophy is to provide rational grounds for preferring one moral theory over another as a basis for making moral judgments. This is not the only function which justification may serve; establishing moral truths and building consensus around moral values are alternatives on either side of this conception. But it is the sense of justification with which I will be concerned here. This is a moderate conception which is identifiably epistemic and yet not committed to the existence of moral truths. The central problem of moral epistemology on this view is to provide an account of the rational grounds on the basis of which one may discriminate between competing theories. Two extreme positions are possible here: one which holds that a method of justification exists for choosing between competitors which depends only upon the nature of rationality (Kant’s view); and another which holds that there are no rational grounds at all for choosing between competing and equally consistent moral systems (moral skepticism).¹

In what follows I will be concerned with an idea about justification which lies between these two extremes, and which is most fully developed in the writings of John Rawls.² This is the method of reflective equilibrium: the idea

¹ Kant provides an example of a highly objectivist moral epistemology: “[Everyone] must concede that the ground of obligation here must not be sought in the nature of man or in the circumstances in which he is placed, but sought a priori solely in the concepts of pure reason.” (Foundations of the Metaphysics of Morals (Bobbs-Merrill, 1959, p. 5) “From what has been said it is clear that all moral concepts have their seat and origin entirely a priori in reason.” (ibid., p. 28) “[Moral laws] are valid as laws only insofar as they can be seen to have an a priori basis and to be necessary.” (The Metaphysical Principles of Virtue (Bobbs-Merrill, 1964), p. 14) Rawls’s recent Dewey lectures (“Kantian Constructivism in Moral Theory,” Journal of Philosophy 77:9 (1980)) present a view which depends upon a consensus-building conception of justification, which will be discussed briefly in the final paragraph of this article.
that moral theories are justified through a process of deliberation in which we consider a wide range of beliefs and judgments as a system, and evaluate competing moral theories in terms of their coherence with this system. The attraction of this model is that it avoids the strong Kantian claim that moral theories may be established unconditionally, but also seeks to avoid moral skepticism. My purpose in what follows is simply to assess the degree to which arguments of this sort succeed in permitting rational choice between theories. I will conclude that the method Rawls describes is in fact very weak, since it depends heavily on the substantive moral presuppositions with which one begins. Instead of justifying a moral theory it seems largely to be a technique for systematizing and organizing one’s antecedent moral convictions. Given that Rawls’s system is one of the most methodologically sophisticated to appear in recent years, this conclusion casts a substantial shadow over the project of providing an objective justification for moral theories in any form.

Reflective equilibrium and coherence arguments. The central idea at work here is the conception of moral philosophy as a process of seeking “reflective equilibrium”: that is, a process through which the philosopher attempts to arrive at a coherent moral theory which is compatible with a wide range of pre-theoretical commitments. This process of justification requires a continual interaction between the formulation of theoretical principles and a consideration of pre-theoretical commitments. We formulate a general moral principle, then test it against our pre-theoretical commitments. Do its concrete consequences correspond to the judgments we would actually make? Does it satisfy conditions which we pre-theoretically believe to validly characterize all acceptable moral theories? If the theory conflicts at either stage with our moral convictions, then we have to decide whether to modify the theory or abandon the convictions. This process is to continue until theory and commitments are in harmony with each other. Rawls characterizes this final stage as the condition of reflective equilibrium. We may refer to this process as one of “reflective deliberation.”

This conception postulates that the preferred moral theory is the one which best serves to organize and explain “considered judgements” and moral convictions. More specifically, Rawls’s view is that we have a whole range of moral commitments, from the level of concrete judgments like “it is unjust for one person to hold all land as private property,” to formal judgements like “moral theories ought to be general, universal, and public” to substantive principles like “it is fair to submit people to principles they themselves would have chosen.” These commitments are substantially pre-theoretical, in that everyone (not just moral philosophers) shares them, or analogous ones. Accordingly, a moral theory is justified to the degree that it serves to
systematize and explicate the full range of these moral commitments, and to extend our commitments in cases where our pre-theoretical judgment is uncertain.

Rawls uses this process of reflective deliberation to support his formulation of the premises of justice as fairness: the idea that the correct theory of justice is that which rational persons would choose to regulate their public affairs, and the particular description which Rawls puts forward of the circumstances of choice (the original position). Having arrived at these ideas, Rawls argues that they strongly support the two principles of justice over other possible candidates. But this argument from the original position to the two principles of justice is intended chiefly as a vivid model of the moral ideas we bring to the problem of justice; it is chiefly intended to provide a way of coherently expressing and extending those ideas.\(^3\) Thus the deductive argument from the description of the original position is not itself a primary source of justification for Rawl’s theory of justice. It depends wholly on the deliberative justification available for the general program of justice as fairness.\(^4\)

The process of reflective deliberation described here depends heavily on a coherence epistemology: the idea that justification depends upon the power of a theory to smoothly incorporate a wide range of diverse moral commitments into a coherent moral conception.

A conception of justice cannot be deduced from self-evident premises or conditions on principles; instead, its justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view. (TJ, p. 21).

---

\(^3\) “One should not be misled, then, by the somewhat unusual conditions which characterize the original position. The idea here is simply to make vivid to ourselves the restrictions that it seems reasonable to impose on arguments for principles of justice, and therefore on these principles themselves.” (TJ, p. 18)

\(^4\) In treating Rawls’s theory of justification I will not discuss his use of rational choice theory. This reasoning might seem to justify the theory of justice directly by showing that it is “the only choice consistent with the full description of the original position.” Rawls describes this style of argument as “a kind of moral geometry.” (TJ, p. 121) However, this application of rational choice theory is embedded within the assumptions defining the original position. These assumptions must themselves be justified on moral grounds, and their justification depends upon the method of reflective deliberation. It thus seems fair to conclude that the use of rational choice theory plays only a minor role in the justificatory argument.
Is there any specifically moral import to coherence arguments? Some philosophers have argued that the coherence methodology entails a sharp distinction between ethics and empirical knowledge. For example, Ronald Dworkin argues that the use of coherence arguments entails that we must accept a constructivist view of moral beliefs, according to which moral truths are *constituted* by the persons who subscribe to them.⁵ He reaches this conclusion through the observation that coherence considerations lead us sometimes to revise basic judgements, sometimes moral principles; and he asks how this could be justified if there were independent moral truths. This argument is unsound, however; for a coherence epistemology in empirical science entails just the same revisability of natural law statements, without entailing that observers constitute those laws. The coherence epistemology is a method of justification, not a thesis concerning the ontological status of the laws arrived at. Consequently, this approach does not entail constructivism.

David Lyons doubts the justificatory import of the coherence approach in “Nature and Soundness of the Contract and Coherence Arguments”.⁶ He holds that since the coherence approach depends upon the considered judgments of the moral philosopher, the principles it supports are no stronger than ordinary moral common sense. “One may still doubt whether a coherence argument says anything about the validity of such principles; for pure coherence arguments seem to move us in a circle, between our current attitudes and the principles they supposedly manifest.”⁷ On Lyon’s view, therefore, the coherence argument serves merely to demonstrate the presuppositions of common morality, and not at all to demonstrate the correctness of the principles which it discovers. However, Lyons’ argument is too broad-gauge; if it were correct it would apply equally to the use of coherence arguments in empirical science as well. But Quine, Goodman, and others have made out a strong case for the viability of a non-foundationalist epistemology.⁸ Consequently, if a coherence-based

---

⁷ *Reading Rawls*, p. 146.
argument is defective in moral reasoning, its defect must derive from the content of the argument and not the general epistemology on which it is based.  

Against these views, I hold that the coherence epistemology is uncontroversial from the moral point of view. It is a thesis about the nature of rational theory-assessment generally, and is as well suited to moral as to empirical reasoning. (This is not to say that the coherence method is uncontroversial, but only that the arguments which may be offered against it are largely independent of the particular features of moral reasoning.) The chief import of Rawls’s choice of the coherence epistemology is its implicit rejection of a foundationalist theory of knowledge, according to which some class of statements is epistemically basic and unrevisionable. Rawls instead borrows a chapter from Goodman and Quine and depicts moral reasoning as a mutual adjustment of beliefs at every level, in which no belief is immune from revision. This position rejects the model which represents knowledge as being founded on some set of privileged statements (perception reports or considered moral judgements), and instead represents the acquisition of knowledge as a process in which theoretical beliefs sometimes dictate the rejection of more immediate judgments. In the moral case, this stance entails rejecting rational intuitionism, or the view that moral theories may be founded on a privileged set moral intuitions. This stance is thus one shared with general epistemology.

If this treatment is correct, then the coherence epistemology ought not be the center of discussion in considerations of Rawls’s theory of justification; rather, the central question is the strength of the theories and judgments which are the content of the process of reflective deliberation. For it must be emphasized that the coherence argument does not itself confer justificatory weight on its conclusion; it is a general form of argument rather than a source of justification. In order to know whether the argument has justificatory weight, it is necessary to know what the content of the argument is. Merely to conclude that “T encompasses all our beliefs in the smoothest and most consistent way” does not serve to justify T. Rather, we must add that the bulk of our beliefs are credible to some degree.

---

9 I will reach something like Lyons’ conclusion below; but the reasoning will not depend on weaknesses of the coherence epistemology generally, but rather the uncertainty of the chief forms of belief which constitute the content of such arguments in moral philosophy.

In order to evaluate Rawls’s justificatory arguments, therefore, it is necessary to consider the strength of the beliefs and judgments which serve as the source of credibility within the process of deliberation described here. This process of deliberation may be carried out at two levels of generality, depending on the sorts of beliefs which are introduced into the process. First (in what Rawls calls “narrow reflective equilibrium”) the process may take account only the “considered judgments” with which the philosopher begins: the concrete moral judgments in which we have the greatest degree of antecedent confidence. Here the task of deliberation is to construct a moral theory which organizes and explains these considered judgments in the most coherent fashion possible. Second, the process may take account of a wide range of philosophical and empirical beliefs as well as considered judgments. This version of the process (which Rawls describes as one pursuing “wide reflective equilibrium”) is much more general than the former, in that it attempts to link the moral theory eventually produced to the full range of moral and empirical commitments with which we begin.

A quasi-inductive argument. In his conception of narrow reflective equilibrium, Rawls proposes that it is possible to “test” a moral theory in terms of its consonance with our antecedent moral convictions (in analogy with the testing of an empirical theory against its consequences for experience). That is, we may evaluate a moral theory by working it out in detail and seeing the degree to which its distant implications match the judgments we are inclined to make in specific cases. “There is, however, another side to justifying a particular description of the original position. This is to see if the principles which would be chosen match our considered convictions of justice or extend them in an acceptable way.” (TJ, p. 19) More fully, Rawls postulates that we all have a psychologically real “moral sense” or moral competence, which leads us to make moral judgments in a variety of actual cases. These judgments form the range of relatively fixed points to which moral theory must generally conform. Any moral theory which widely and grossly disagrees with the spectrum of our considered judgments must be judged inadequate; our initial moral opinions are presumed to have some preliminary standing to which the moral theory must conform.

An especially crisp formulation of this idea occurs in “The Independence of Moral Theory”: “People profess and appear to be influenced by moral conceptions. These conceptions themselves can be made a focus of study; so provisionally we may bracket the problem of moral truth and turn to moral theory: we investigate the substantive moral conceptions that people hold, or
would hold, under suitably defined conditions.”

A similar view is expressed in A Theory of Justice: “Now one may think of moral philosophy at first. . . as the attempt to describe our moral capacity; or in the present case, one may regard a theory of justice as describing our sense of justice. . . “ (TJ, p. 4)

This model of moral philosophy borrows from Noam Chomsky’s description of the nature of linguistic theory. Its premise is certainly correct: we do in fact make a vast number of moral judgments without being explicitly aware of their relation to more general moral principles. So the project of constructing a moral conception which best organizes and supports these judgments is a reasonable one. Rawls is also correct in supposing that theories of this sort may only be evaluated on the basis of their success at organizing the “behaviors” to which they correspond: the actual moral judgment-making practices which the theory is intended to account for. Through the process of deliberation which Rawls describe, we may arrive at a theory of the moral competence of the subject: a set of moral principles which best represents S’s preliminary considered judgments, in much the way that a linguist arrives at a generative grammar to represent judgments of grammaticality. Thus it is reasonable to say that a moral theory may in some sense be “tested” by examining its fit with actual practices.

However, this linguistic analogy by itself does not allow us to describe this process as one supplying moral justification. Rather, this analogy suggests that the process described here is a purely empirical one: the theory which emerges is a theory of a psychological capacity (the capacity to make moral judgments), and is true or false as a description of the psychological makeup of the agent. But nothing in this process as described up to this point provides any moral justification for believing that the resulting theory is correct: the fact that T best systematizes S’s moral judgments does not serve to show that T is morally...

---

12 “We thus make a fundamental distinction between competence (the speaker-hearer’s knowledge of his language) and performance (the actual use of language in concrete situations).... The problem for the linguist ... is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer. . . . A grammar of a language purports to be a description of the ideal speaker-hearer’s intrinsic competence.” Noam Chomsky, Aspects of the Theory of Syntax (Cambridge: MIT, 1965), p. 4.
correct. Thus the linguistic analogy represents moral philosophy only as a division of psychological theory.

This process of deliberation might seem to provide a basis for moral justification, however, if we shift metaphors and conceive of it as a form of moral “induction”. The fact that “T is the best available theory of S’s moral competence” would gain standing as a moral justification if we were to suppose that S’s moral judgments are usually morally correct. On this account, considered moral judgments are analogous to observation reports in empirical science. They have an initial moral credibility which can be transferred to the theories which moral philosophers construct. If we make this assumption (analogous to the assumption in science that our perceptions are generally veridical) then the following quasi-inductive argument has weight:

i. S’s considered judgments are by and large morally correct.
ii. The principles of justice best explicate those considered judgments.
iii. Therefore the principles of justice are at least approximately correct.

This arguments represents a case in which a class of judgments is presumed to have some *prima facie* credibility, and that credibility is passed on to the theory which best explains the judgments.

This interpretation of the moral competence model of justification depends upon a strong analogy between considered judgments and observation reports in empirical science. The antecedent credibility of considered judgments is transmitted to the moral theory. But the disanalogies between these two sorts of beliefs seem more striking than their similarities. We have reasons for being confident in observation reports which do not pertain to considered judgments.

First, our knowledge of the causal basis of perception gives us reason to believe that empirical observations are likely to be true. Using physical theory we can vindicate our reliance on observation reports as the epistemic basis of empirical knowledge. By contrast, considered moral judgments plainly do not have this property. Nothing we know about the psychology of moral belief formation gives us any reason for supposing that the beliefs which emerge from that process are likely to be correct. Rather, accidents of familial, religious, and educational experience seem to have a major influence on the particular constellation of beliefs which eventually express our moral competence.

A second and related point has to do with differences in the degree of theory-dependency characteristic of these two sorts of beliefs. Notwithstanding post-positivist critiques of the observation-theoretic distinction, it is reasonable
to conclude that observation reports in empirical science are theory-dependent to only a small degree: the observations made by holders of different empirical theories are generally relatively invariant. By contrast, considered moral judgments are inseparable from the moral theories which they are supposed to evaluate. This is so because moral judgments are not primitive beliefs; rather, they generally reflect a process of reasoning through which the agent weighs relevant principles and facts in order to arrive at a concrete judgment. The resulting judgment therefore depends heavily upon the background moral theories of the agent.

Third, considered judgments seem to be no more certain than the moral theories which they are intended to evaluate. In fact the existence of widespread disagreement at the level of concrete moral judgments is one of the sources of our impulse to arrive at more general statements of moral principle. In large part we want a justified moral theory in order to resolve disagreements at the level of considered judgments. Thus the analogy between considered judgments and observation reports is unpersuasive in view of the controversial disagreements to be found among the moral competences of different individuals and groups. Whereas observation reports serve to validate or refute empirical theories, concrete moral judgments are themselves often validated or refuted by appeal to more general moral theories.

These considerations suggest that the competence argument is only weakly justificatory in the moral sense defined above. It depends upon the approximate correctness of our moral competence. This appeal would be strongly justificatory only if we could treat considered judgments as relatively credible. But considered judgments are not epistemically similar to perceptual reports; they lack the form of immediacy and reliability which observation reports

---

14 W. H. Newton-Smith gives a sophisticated defense of a rationalistic interpretation of scientific knowledge in his recent *The Rationality of Science* (Boston: Routledge & Kegan Paul, 1981). A crucial part of his argument hinges on his defense of the relative fixity of observation through change of theory. “We must not make the fallacious inference that, as any observation report no matter how well-corroborated may be rejected by appeal to theory, all observation reports might be rejected by appeal to theory. For we can reject particular reports involving O-terms only if we have strong evidence (which will rely on other observational reports) for a theory.... While we cannot have absolute faith in any particular reports or a particular range of reports, we are entitled to have general faith in the low-level O-reports we are inclined to make.” (p. 28) See his detailed criticisms of Kuhn and Feyerabend in chapters 5-7.
possess, in virtue of which they function as an empirical control on physical theory. We therefore have little reason for confidence in the correctness of our considered judgments; but narrow reflective equilibrium makes the resulting moral theory only as strong as the considered judgments themselves. It is thus reasonable to hold that the primary import of the process of narrow reflective deliberation is empirical rather than moral. The theory which results from this form of deliberation is a good description of existing moral competence, but the process provides only a very weak normative basis for evaluating the moral theory against other moral theories. This form of reasoning thus falls under empirical moral psychology rather than normative moral theory.\footnote{This conclusion is in line with Rawls's view that an important part of moral philosophy is the description of alternative moral conceptions. Thus “one may think of moral philosophy at first . . . as the attempt to describe our moral capacity. “ (\textit{TJ}, p. 46)}

\textit{Inter-theoretical support.} Let us turn now to Rawls's concept of “wide reflective equilibrium.” According to this approach, “. . . one is to be presented with all possible descriptions to which one might plausibly conform one’s judgments together with all relevant philosophical arguments for them. “ (\textit{TJ}, p. 49) The correct moral theory is that which would be selected through this highly general form of deliberation. This method requires that we treat our considered judgments, philosophical theories, and proposed moral theories as a system, and modify various parts in order to improve their consistency, until no further changes are required. In this form the considered judgments are open to substantial revision.

In a series of articles Norm Daniels has elaborated on the method of wide reflective equilibrium. He represents wide reflective equilibrium as “an attempt to produce coherence in an ordered triple of sets of beliefs held by a particular person, namely, (a) a set of considered moral judgments, (b) a set of moral principles, and (c) a set of relevant background theories,”\footnote{Norman Daniels, “Wide Reflective Equilibrium and Theory Acceptance in Ethics,” \textit{The Journal of Philosophy} (1979), p. 258. (For a similar treatment of Daniels’ view see also his “Reflective Equilibrium and Archimedean Points,” \textit{Canadian Journal of Philosophy} 10:1 (March 1980).) He maintains that the addition of (c) makes the problem of justification more tractable, insofar as the background theories provide some support for the moral principles which is independent of the considered moral judgments.

On this account, the method of wide reflective deliberation gets its justificatory bite from the fact that we may bring theories from other areas to
bear on the choice between competing moral theories, and those background theories may possess independent credibility. Daniels cites Rawls’s theories of the person, of procedural justice, and of the role of morality in society as examples of the sorts of background theories he has in mind; other examples would include the empirical theories we hold of human psychology, social dynamics, and the like, on the basis of which we consider the likely stability of proposed moral arrangements. Supposing that those theories have some degree of credibility which derives from considerations distinct from those available for the moral theory itself, and that these theories offer grounds for discriminating between competing moral theories, we will be able to justify one moral theory over another in terms of their relative compatibility with the background theories.

This approach represents a different analogy with justification in natural science: in this case, not inductive support, but inter-theoretical support. Moreover, Daniels’ reconstruction is well-suited to providing a form of justification of the sort needed here (to provide reasons for preferring one moral theory to another). If we have reason to believe the background theories and if they support one moral theory over another, then we have reason to accept the favored moral theory. In other words, the method of wide reflective equilibrium seeks to embed disagreements about moral theory within a wider framework in which rational argument may get some traction.

This move represents progress, however, only on the assumption that philosophical theories are somehow easier to justify than moral theories in the first place. And in fact Daniels claims that “disagreements about [background] theories may be more tractable than disagreements about moral judgments and principles. Consequently, if the moral disagreements can be traced to disagreements about theory, greater moral agreement may result.” Two points may be made about this view. First, it may be noted that Rawls seems to reject this move in “The Independence of Moral Theory”, where he maintains that substantive moral questions are largely independent of background philosophical theories (e.g. theories of the person, etc.). “Now my thought is this: much of moral theory is independent from the other parts of philosophy.” And second, a certain amount of skepticism may greet Daniels’ expectation that disagreements over theories of the person or of the role of justice in society are less controversial, or more easily resolved, than disagreements about moral principles. It seems as reasonable to suppose that

---

broadening the discussion from moral judgments and principles to moral judgments, principles, and philosophical theories, has simply broadened the possible sources of irresolvable disagreement.

Let us focus the latter criticism more sharply. To what extent are there strong forms of justification available in support of philosophical theories (e.g. theories of the person)? The justifications offered for such theories are apparently as inconclusive as the justifications offered for moral principles in the first place. Some candidate theories of the person may be ruled out on the grounds of internal inconsistency; but among the surviving class we can only discriminate on grounds of consistency with other radically uncertain beliefs. To make one relevant example: Rawls criticizes utilitarianism on the ground that its associated theory of the person is defective. It treats persons as mere receptacles of happiness rather than morally significant units. “Utilitarianism does not take seriously the distinction between persons.” (TJ, p. 27) But a committed utilitarian may be as little persuaded by Rawls’s theory of the person as he is by the full theory of justice. The arguments Rawls would give to support this background theory seem neither stronger nor weaker than those available for the full theory of justice. In this instance, then, retreat to the background theory of the person represents no justificatory gain.

To broaden the argument in the way that Daniels proposes, therefore, seems not to represent an epistemic gain, or a gain in the rational confidence we may have in the correctness of a given moral theory. This failure derives immediately from the fact that we have no greater warrant in holding typical philosophical theories than we do in holding the moral theories we are trying to justify. What Daniels’ move does allow, however, is a wider basis on which to achieve consensus among particular discussants. If it happens that the philosophical theories appealed to are in fact accepted on all sides of a particular moral debate (whatever their objective warrant), then this move may permit the gaining of consensus where none was possible previously. Consider discussants Smith and Jones who disagree over moral theories T and T’. Suppose T and T’ are equally consistent with all considered judgments shared by Smith and Jones, so both discussants may continue to hold their distinct views without inconsistency. But when we introduce a philosophical theory P which Smith and Jones both accept, it may emerge that P supports T and is inconsistent with T’. In this case Smith and Jones must come to agree on the correctness of T, on pain of inconsistency. Thus appeal to background philosophical theories may permit us to achieve consensus over a moral theory; but (unless we have independent reasons for believing these theories to be correct) such an appeal does not increase the rational warrant we may attach to the moral theory.
Let us briefly consider Daniels’ idea that the background theories in question should have independent credibility. How shall we define the requirement of independence here? Daniels chooses an analogy with natural science, in the distinction between accidental generalizations and laws of nature. His criterion is that “we should require that the background theories in (c) be more than reformulations of the same set of considered moral judgments. . . . The background theories should have a scope reaching beyond the range of the considered moral judgments used to test the moral principles.” He thus construes the background theories in analogy with theories of natural science: each is tested by reference to concrete judgments (experimental findings and considered moral judgments), and an accidental generalization is distinguished from a real law in that the latter has test circumstances which are disjoint from the class it is advanced to encompass. In the moral case, this criterion would require that the background theories be supported by a set of considered moral judgments disjoint from those currently under consideration (referred to as (a) in Daniels’ formulation). Daniels asserts that Rawls’s background theories are independent in the sense described here. “His central level III theories of the person and of the role of morality in society are probably not just recharacterizations or systematizations of level I moral judgments.” This criterion seems to threaten collapse of Daniels’ three-part account of wide reflective equilibrium, however. The class (a) ought ideally to include all available considered moral judgments, in which case the class constraining (b) is identical with that constraining (c). In this case there is no gain in justificatory force.

These remarks suggest that the method of wide reflective deliberation does little to enhance the justificatory resources available to moral theory. Rather, it relates the problem of justifying moral theories to that of justifying philosophical theories. This is at best a lateral move, however, since philosophical theories are themselves notoriously controversial, and the question of justification of such theories is equally unresolved. To what extent, then, do arguments based on the method of wide reflective equilibrium serve to justify particular moral theories (that is, to provide rational grounds for preferring one theory over another)? Only to the degree that all discussants share (or can be brought to share) the philosophical theories appealed to in the argument. The justification offered here is relative to a set of background

---

assumptions, with the result that if these are not satisfied then the argument has no justificatory grip.

Assessment. Let us take stock. To what extent does the process of reflective deliberation provide a basis for comparative evaluation of alternative moral theories? There are three cases to consider. First, against a given set of considered judgments and background theories, two moral theories may be evaluated in relation to their respective abilities to encompass these beliefs. Here the coherence argument has some bite, since it permits a ranking of competing theories in terms of their fit with background beliefs. In the case in which all discussants share a significant number of considered judgments and background theories, the method of reflective deliberation can have justificatory force; it can provide reasons for the discussants to prefer one theory over another.

Second, it may be that two competing moral theories are equally consistent with the background beliefs, but inconsistent with one another. In this case the method ranks them equally and provides no basis for choosing between them. This possibility seems to be part and parcel of the coherence epistemology. This is Quine's point about the underdetermination of physical theory by empirical evidence. Given an incoherence within a theoretical system, there will generally be alternative and non-equivalent ways of resolving the contradiction. So we may suppose that the process of deliberation might have led to a family of non-equivalent moral theories. Each is a coherent system given the totality of background beliefs, and consequently the coherence epistemology offers no basis for choosing between them. This observation suggests a strong conclusion: indeterminacy of moral theory given all possible background beliefs. If this is a real possibility, then it is possible for all

---

21 Quine holds that theoretical beliefs are in principle underdetermined by empirical evidence. “If by some oracle the physicist could identify outright all the truths that can be said in common-sense terms about ordinary things, still his separation of statements about molecules into true and false would remain largely unsettled.... Conceivably the truths about molecules are only partially determined by any ideal organon of scientific method plus all the truths that can tee said in common-sense terms about ordinary things.” Word and Object, pp. 21-22. This underdetermination follows from the fact that it is possible to adjust a set of theoretical beliefs in non-equivalent ways in order to bring the system into harmony with experience. “The total field [of beliefs] is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience.” (“Two Dogmas of Empiricism,” p. 42.)
discussants to share the bulk of their background beliefs and considered judgments, but still disagree irresolvably about moral theory. This possibility undercuts the hope of finding a uniquely best moral theory, and hence the idea of moral truth.

Third, the method provides no basis at all for comparing moral theories deriving from different sets of background beliefs. If two persons have irresolvable disagreements at the level of their background beliefs, then each will find a reflective equilibrium around different moral theories, and the method will be unable to rank these alternatives. This possibility suggests that Rawls’s system of argument has strongly relativist implications: the correct moral theory depends upon the background beliefs with which one begins, and these may differ in important ways from one community to another.

Thus the method of reflective deliberation is a weak form of justification at best. This is so because the requirement of consistency upon which it depends is itself a weak constraint, and because the constraining beliefs involved in the process are themselves highly uncertain. We may well imagine that there are alternative moral theories which are equally good on coherence grounds as reconstructions of a fixed set of background beliefs. And among these alternatives the method of reflective deliberation provides no ranking whatsoever. More radically, the method gives no way to begin to resolve moral disagreements among discussants having substantially different starting beliefs and judgments.

We concluded above that the coherence epistemology itself is equally appropriate in ethics and science. We have now seen, however, that there is a crucial difference in the epistemic standing of science and moral philosophy. This difference derives from the availability of a source of independent control in the coherence argument. A chief part of the objectivity of science stems from the relative independence of perception from background theory. When we add this feature to the process of seeking a coherent system of belief, we get a process of knowledge-formation which is arguably objective and truth-attaining. But arguments offered above suggest that there are no analogous controls in moral reasoning. Considered judgments and general philosophical theories seem like the only available candidates for such an independent control; but neither will perform that function. Considered judgments cannot perform this function because they derive from our background moral theory. And background philosophical theories are even less suited to function as an independent control on theory formation, since they are themselves only weakly controlled by rational argument.
There is thus a crucial disanalogy between the two areas: empirical theory is constrained by evidence which is in a rough and ready sense independent of the theoretical conclusions we eventually arrive at, whereas moral deliberation is not. But if there is no class of beliefs which have an antecedent credibility independent of their relations to other statements, then how does any epistemic warrant enter the system of moral deliberation? We must conclude that the method of reflective deliberation provides only a very weak scheme of justification; it depends heavily and unavoidably on the particular considered judgments and background beliefs with which we begin, and these are themselves unjustified. The most this process can do is to bring to the surface the principles which underlie the moral practices of the existing community, and subject those practices to a limited form of rational criticism.

This low estimate of the epistemic force of the method of reflective deliberation may help to explain what is otherwise an extraordinary fact about Rawls’s “Kantian Constructivism in Moral Theory”, the fact that this essay turns sharply away from the idea of moral epistemology. In these lectures Rawls maintains that “the ‘real task’ of justifying a conception of justice is not primarily an epistemological problem. The search for reasonable grounds for reaching agreement rooted in our conception of ourselves and in our relation to society replaces the search for truth.” For the process of reflective deliberation we have analyzed here provides a much better basis for the problem of consensus-building than it does for establishing correctness. The fact that the process of deliberation depends upon shared background beliefs has an obvious political counterpart: it will only be possible to build a democratic society if there is a substantial degree of agreement about basic values and principles. Within the context of such agreement, the process of reflective deliberation provides an attractive model of a process through which agreement might be reached.

\[\text{22} \text{ Rawls, “Kantian Constructivism in Moral Theory,” pp. 515-572.}\]
\[\text{23} \text{ Rawls, “Kantian Constructivism in Moral Theory,” pp. 518-19.}\]
\[\text{24} \text{ An earlier version of this paper was presented to the New York State Philosophical Association (the Creighton Club) in October, 1981. I gratefully acknowledge the useful criticisms received on that occasion.}\]