

Rawlsian Relativism and Background Theories

DAVID WEISSBORD

University of North Carolina at Chapel Hill

Throughout John Rawls' A Theory of Justice (TJ)¹ there is a tacit, but strong, reliance upon a notion of intuition that was first developed in an earlier paper entitled, "Outline of a Decision Procedure for Ethics."² It is the role of intuition in a decision procedure for ethics and, specifically, in the method of reflective equilibrium that is attacked by Peter Singer, R. M. Hare, and Richard Brandt. Attacks that, if successful, would reduce the Rawlsian scheme to an undesirable form of relativism. This relativism reduces to an internal test of coherence, where there are no "external" principles to which we can appeal, no "external" standards by which we can judge the principles and intuitions themselves. Singer questions why the moral judgments we intuitively make are legitimate fixed points against which theories can be tested. Why not, he asks, make the opposite assumption, i.e., that such judgments are likely "to derive from discarded religious systems, from warped views of sex and bodily functions, or from customs necessary for the survival of the group of social and economic circumstances that now lie in the distant past?" Hare claims that Rawls advocates a narrow subjectivism that is the result of grounding his theoretical structure on "a cosy unanimity" in people's considered judgments.³

After examining the charges made by Singer, Brandt, and Hare, we will turn to two attempts to defuse those charges. The first is that of Norman Daniels, who concentrates his response on wide reflective equilibrium and its flexibility in extensively revising moral judgments. Seeing where Daniels fails is instructive in examining the second attempt made by Ronald Dworkin. Dworkin constructs his defense by analyzing Rawls' characterization of our capacity to make moral judgments about justice. This analysis leads to Dworkin's "constructive" model which purportedly explains why coherence is required. Dworkin's account is instructive in helping us see why it is critical to get clear about the nature of the subject to whom principles of justice apply. It is in light of these problems that "background theories" emerge, which are seen to be instrumental in the Rawlsian system and, as such, un-

wise to ignore when making the kinds of claims made by Singer, Brandt, and Hare.

I

In his "Sidgwick and Reflective Equilibrium," Peter Singer relies on A Theory of Justice in evaluating Rawls' method of reflective equilibrium. Singer's charge is that Rawls is a moral subjectivist. He claims that this "follows from (Rawls') views that the validity of a moral theory will vary according to whose considered moral judgments the theory is tested against." This claim is based on Singer's belief that the moral theories produced by the procedure have no room for a validity that is independent of achieving reflective equilibrium. Singer concludes, after citing passages in TJ that define and explicate reflective equilibrium, that our moral capacity is merely described and that the description must be dependent for its correctness on the way it fits what it describes. If this is accurate, Rawls would be committed to the view that there is no meaningful sense of 'valid' that goes beyond the achievement of a stable reflective equilibrium. There is "nothing to suggest room for a notion of validity beyond the conjunction of theory and revised moral capacity" (SRE, 494).

Singer pushes his claim by making the following suggestion: "If I live in one society, and accept one set of considered judgments, while you live in another society and hold a quite different set, very different moral theories may be 'valid' for each of us. There will then be no sense in which one of us is wrong and the other right" (SRE, 494). Singer is partially correct here, insofar as the Rawlsian model does admit of different considered moral judgments given various background conditions and therefore, different states of reflective equilibrium.⁴ But Singer fails to consider seriously the very basis of the original position. Given our nature as free and equal rational beings presented with a limited class of facts that express and are based upon that nature, we think of ourselves in a certain light in the original position that is almost constant and, for Rawls, therefore, close to objective. Singer's hypothetical begins to have less force in virtue of the constraints placed on individuals in the original position, especially insofar as it is impossible to tailor principles to the circumstances of one's own case.

II

Richard Brandt, in his A Theory of the Good and the Right, discusses the role intuitions have historically played in ethical theory and, particularly, in Rawls' A Theory of Justice.⁷ His forthcoming attack is presaged

in the opening pages, as he remarks that intuitionism suggests that we already "have presumably well-justified opinions about the answers to the several traditional questions . . . although these opinions need to be systematized and hence, in some cases, revised to some degree . . . we roughly already know most of the answers" (TGR, 3). In reference to Rawls' claim that our conjectured principles can be checked against our considered judgments in reflective equilibrium (TJ, 51), Brandt queries: "What kind of thing do these philosophers have in mind when they speak of 'intuitions' against which normative generalizations can be checked?" Brandt discounts self-evident truths and special "senses" as resting at the bottom of the views of contemporary philosophers. Having set the stage, by discussing and rejecting other views of intuition, Brandt transits directly into the Rawlsian account.

Brandt begins by relating the Rawlsian model to the procedure Nelson Goodman recommended for logic. If we are unwilling to accept particular implications yielded from a general principle, we ought to amend that principle; unwilling to amend a general principle, we ought to reject a particular judgment if it violates that principle. In general, when we are faced with having to choose between two sets of normative beliefs, we ought to give preference to that set we find easier to believe as a whole. This is what Brandt refers to as a 'coherence theory of justification in ethics' (TGR, 19). Brandt then notes how Rawls qualifies the Goodman coherence view by explaining the difference between reflective equilibrium 'in the narrow sense' and 'in the wide sense'. Narrow reflective equilibrium "contemplates a Goodman-like combination of restructured moral principles of persons . . . who have been uninfluenced by philosophical arguments; the latter (wide reflective equilibrium) contemplates a Goodman-like system of intuitions as these are after influence by philosophical arguments . . ." (TGR, 19).

At this point, Brandt explains that the most problematic aspect of such a coherentist view of ethics is that we are merely required to hold a consistent set of judgments about types of action, without dictating "any particular way of adjudicating conflicts between, say, higher-level and lower-level normative 'intuitions'" (TGR, 20). Without any way of adjudicating between conflicts, proponents of such a coherence-intuitionist theory "must find some other justification for . . . giving preference to the set of normative beliefs we find . . . easier to accept as a whole . . . The fact that a person has a firm normative conviction gives that belief a status no better than fiction. Is one coherent set of fictions supposed to be better than another?" (TGR, 20).

Brandt asks why we should desire wide reflective equilibrium over narrow reflective equilibrium. "Is it

thought that somehow we thereby approach sophistication?" (TGR, 21). For him, it is puzzling that an intuition should be a test of anything, since facts about the genesis of our moral beliefs (viz. that normative beliefs are largely a function of the particular tradition which nurtured us) "militate against mere appeal to intuitions in ethics" (TGR, 21). He claims:

What we should aim to do is to step outside our own tradition somehow, see it from the outside, and evaluate it, separating what is only the vestige of a possible once useful moral tradition from what is justifiable at present. The method of intuitions in principle prohibits our doing this. It is only an internal test of coherence, what may be no more than a reshuffling of moral prejudices . . . the search for reflective equilibrium will only produce different moral systems, and offers no way to relieve . . . conflict(s). (TGR, 21-22)

According to Brandt, then, the goal appears to be one of making a coherent system of intuitions. But is that all that is involved? Is such coherence a necessary and sufficient condition for a fully justified system containing normative claims that are related to our actions? A closer look at Rawls, later in this paper, will attempt to show that such coherence is only a necessary, not a necessary and sufficient, condition for that justified system.

Finally, Brandt duly notes that Rawls considers a virtue of his conceptual framework the notion of everyone making the same choice of a moral system, so that what is right is so for everyone. But Brandt asks incredulously: "What is the force of knowing that one would, along with all others, have chosen certain moral principles if one had been making a choice behind a veil of ignorance, in the 'original position'?" (TGR, 243). Brandt suggests that this sort of information would not justify the chosen moral code, in the sense of motivating a person to support it (TGR, 243). This isn't clear. Does Brandt mean here that a moral code is justified insofar as one is psychologically motivated to support it or insofar as one is justifiably motivated? The latter is obviously circular and the former seems almost trivial. For if I am not psychologically capable of supporting what anyone in the original position would consider just, then that sense of justice must be modified to accord with the psychological dispositions of, and reasonable demands that can be made upon, people for whom the moral code is applicable. Further, Brandt suggests that there is no serious defect if fully rational persons do not support the same moral system (TGR, 242). But if we eliminate the initial situation with all of its constraints set

by Rawls, as Brandt intimates we could, we would be left with a moral system that, even if we could get unanimous agreement concerning basic moral principles, would be nothing more than a mere consensus of opinion. Indeed, if this consensus were reached in reflective equilibrium, we would still need some check, namely, a reference to those principles chosen in the original position. Rawls, then, needs to supply an account of how the original position can provide us with an independent check on our principles and considered convictions. Brandt does not think such an account is forthcoming and, as such, argues that Rawls remains mired in cozy coherence.

III

In "Rawls' Theory of Justice," R. M. Hare argues that the normative force in TJ is unsupported by any firm arguments, since Rawls 1) has misconceptions concerning philosophical methodology, and 2) doesn't pay sufficient attention to ethical analysis, and as a result of 1 and 2 lacks the equipment necessary to handle moral methodology (RTJ, 81). Basically, the claim is that Rawls advocates "a kind of subjectivism." In investigating this claim, it will be necessary to explore some of the critical contentions in Hare's thesis and expose his misunderstanding of the Rawlsian project and terminology.

In citing a question raised by Plato, Hare believes he captures the heart of the alleged Rawlsian subjectivist behemoth.

If a man starts from something he knows not, and the end and the middle of his argument are tangled together out of what he knows not, how can such a mere consensus ever turn into knowledge? (Republic 533c) (RTJ, 84)

Like Brandt, Hare takes issue with the idea that a mere consensus approaches knowledge. After quoting Rawls ("For the purposes of this book, the views of the reader and the author are the only ones that count" TJ, 50), Hare lambasts that view, charging that so long as Rawls finds many people with whom he shares "a cosy unanimity in their considered judgments, he and they will think that they adequately represent 'people generally,' and congratulate themselves on having attained the truth" (RTJ, 82). For Hare, the "only 'moral' theories that can be checked against people's actual moral judgments are anthropological theories about what, in general, people think one ought to do, not moral principles about what one ought to do" (RTJ, 86).

For Hare, the Rawlsian project reduces to the answer to this question: "Am I right in what I say about moral questions?" Which, in turn, consists in the an-

swer to the question: "Do you, the reader, and I agree in what we say?" (RTJ, 82). Hare's point is that no claim to justification is made except relative to the acceptance of the considered judgments by readers who share Rawls' intuitions. In light of this discussion, Hare notes despairingly that Rawls claims objectivity for his principles.⁹

Hare's main contention is that, although Rawls is a "self-styled" objectivist, he falls deeply into subjectivism. The subjectivism Hare has in mind relates to Rawls' claim that he can "check" his theory against his and other people's views so that a disagreement between the theory and the views tells against the theory" (RTJ, 83). Hence, reasons Hare, the "truth of the theory depend(s) on agreement with people's opinions" (RTJ, 83). Also, it is Rawls' version of intuitionism that irks Hare. On Rawls' own account, "there is no reason to suppose that we can avoid all appeals to intuition The practical aim is to reach a reasonably reliable agreement in judgment in order to provide a common conception of justice" (TJ 44, 124). A major problem with this view, claims Hare, is that the "theoretical structure is tailored at every point to fit Rawls' intuitions;" hence, it is not surprising "that its normative consequences fit them too--if they did not, he would alter the theory" (RTJ, 84; refers to TJ, 19, 144).

Finally, Hare predicts Rawls' response: "Any justification of principles must proceed from some consensus" (RTJ, 84; TJ, 581). Hare's riposte: justification should end in consensus as a result of argument. Prior consensus is necessary concerning matters of fact, interests of the parties, matters of logic, but not on substantial moral questions.

IV

In his paper "The Original Position" Ronald Dworkin examines two models, one of which, he claims, is presupposed by Rawls' methodology in TJ.¹⁰ Since, according to Rawls, the conditions embodied in the original position are the fundamental "principles governing our moral powers, or, more specifically, our sense of justice" (TJ, 51), Dworkin concludes that Rawls "wants to characterize the structure of our (or, at least, one person's) capacity to make moral judgments of a certain sort, that is, judgments about justice . . . the original position is therefore a schematic representation of a particular mental process of at least some, and perhaps most, human beings" (OP, 25). This suggests to Dworkin that the original position is an "intermediate conclusion, a halfway point in a deeper theory that provides philosophical arguments for its conditions" (OP, 26). In attempting to expose that deeper theory, Dworkin discusses two models, only one of which can

rest at the foundation of the deeper theory. For our purposes, it will be important to determine whether the model he claims to be constitutive of our moral capacity in the Rawlsian system can withstand attacks made on that system by Hare, Singer, and Brandt.

Dworkin begins his exposition by explicating reflective equilibrium. He then states that equilibrium supposes a coherence theory of morality. In explaining why coherence is required, he describes his two models, one of which, it must be remembered, supports Rawls' deeper theory and which, it is now suggested, supports the equilibrium technique. In what follows, I shall sketch the two models, paying particular attention to the "constructive" model which is favored by Dworkin, and then raise questions concerning it and its efficacy in helping Rawls escape the charge of subjectivism.

The first is the "natural" model. It holds that theories of justice describe objective moral reality; principles of justice are innate categories of morality common to all men, imprinted in their neural structure, so that man could not deny these principles short of abandoning the power to reason about morality at all" (OP, 26). Principles of justice, on this view, are not created but are discovered. The imprints in the neural structure produce concrete intuitions which, in turn, "are clues to nature and existence or more abstract and fundamental moral principles . . ." (OP, 28). The individual, then, assumes his moral intuitions to be discrete observations of moral reality.

There are at least two obvious problems with this model. First, adherents to the natural model must suppose that intuitions override explanation. In cases where an individual's intuitions, as well as intuitions in a group, conflict, there is a faith that some explanation exists, although such an explanation is not at hand. Second, it cannot explain, in the Rawlsian scheme, why we are justified in amending intuitions to derive a more secure "fit" (OP, 27-37ff).

The model which Dworkin subsequently adopts is the "constructive" model. It is this model that purportedly explains why coherence is required and supports the equilibrium technique. He holds that intuitions do not suggest the existence of independent principles, but rather are "stipulated features of a general theory to be constructed" (OP, 28). On this view, the judgments which serve as the basis for action must be coherent so that the program of action can be coherent as well. We are not to submerge inconsistencies in our intuitions, accepting on faith that there is an explanation "out there" to be, perhaps, discovered in time; we are, instead, to act on principle and integrate our intuitions into a coherent whole.

The advantage of this model is that it can be applied to a large body of people suggesting a clearly articulated public program essential to any conception

of justice (OP, 30). Hence, it is a "theory of community rather than of particular individuals" which is important for, among other things, adjudication. It allows us to accommodate the community's common convictions "with no claim to a description of an objective moral universe" (OP, 31).

The predictable charge of subjectivism now raises its voice, cutting to the very core of Dworkin's "constructive" model. Dworkin, not oblivious to these concerns, responds by noting that this model "assumes that the men and women who reason within the model will hold sincerely the convictions they bring to it, and that this sincerity will extend to criticizing as unjust political acts or systems that offend the most profound of these" (OP, 30). But this will not do; Dworkin needs to show more than the fact that people truly believe in their convictions. His psychological point is trivially true and applies to the natural model as well. Aware of this last fact, Dworkin tries to cut a slice of both models and merge them into his own pie. "The (constructive) model does not deny, any more than it affirms, the objective standing of any of these convictions; it is therefore consistent with, though as a model of reasoning it does not require, the moral ontology that the natural model presupposes" (OP, 30). This leaves open the possibility that the constructive model can capture moral truth; the trick is that it doesn't require it. This claim is both uninteresting and begs the question. It is uninteresting since other models might just happen to be consistent with the moral ontology required by the natural model, in that within these models we might arrive at moral truth. It begs the question, since we need to be shown why we have a good reason to suppose that the constructive model is, in fact, consistent with the natural model's moral ontology. Without an argument, Dworkin supposes what needs to be demonstrated. Talk about sincere convictions is no demonstration.

V

Norman Daniels argues that Rawls' theory does not grant privileged epistemological status to initial moral judgments or intuitions or, for that matter, to moral principles; nor is his theory a simple coherence view of justification, whereby mere coherence between principles and judgments gives our moral arguments justificatory force. In "Wide Reflective Equilibrium and Theory Acceptance in Ethics," Daniels contends that Rawls' method of wide reflective equilibrium 1) reveals a greater complexity in the structure of moral theories than the traditional view and, as a result 2) may render theory acceptance in ethics a more tractable problem and, if so, 3) may permit us "to recast and resolve some traditional worries about objectivity in

ethics."¹¹ Daniels' self-imposed task is to defend reflective equilibrium against charges that it is a form of moral intuitionism and subjectivism.

Daniels prepares his defense by first explaining what reflective equilibrium is. Roughly, it is the attempt to produce coherence in an ordered triple set of beliefs held by a particular person. That set consists of (a) a set of considered moral judgments, (b) a set of moral principles, and (c) a set of relevant background theories. The best fit between (a) and (b) gives us a narrow equilibrium, which does not provide a basis for a justificational argument, since the initial considered moral judgments do not have any special epistemological status and since the moral principles need only fit with those problematic considered moral judgments. Rawls admits as much in his description of this weaker form of reflective equilibrium:

(When one is) presented with only those descriptions which more or less match one's existing judgments except for minor discrepancies . . . we would be describing a person's sense of justice more or less as it is although allowing for the smoothing out of certain irregularities. (TJ, 49)

This state of affairs is clearly inadequate. Rawls demands more:

(When one is) presented with all possible descriptions to which one might plausibly conform one's judgments together with all relevant philosophical arguments for them . . . a person's sense of justice may or may not undergo a radical shift. Clearly, it is (this) kind of reflective equilibrium that one is concerned with in moral philosophy. (TJ, 49)

Although Rawls admits that this kind of reflective equilibrium is an ideal that we can only approach, in approaching it we depart from narrow reflective equilibrium and "advance philosophical arguments intended to bring out the relative strengths and weaknesses of the alternative sets of principles" (JP, 258). Upon what are these arguments based? It is here that Daniels touches upon the nerve that so many critics gloss over in their attacks on Rawls. Daniels claims that "these arguments can be construed as inferences from some set of relevant background theories" (JP, 258).

The background theories in (c) should show that the moral principles in (b) are more acceptable than alternative principles on grounds to some degree independent of (b)'s match with relevant considered moral judgments in (a). If they are

not in this way independently supported, then there seems to be no gain over the support the principles would have had in a corresponding narrow equilibrium, where there never was any appeal to (c). (JP, 259)

If we can generate acceptable background theories, the principles in (b) would no longer be mere accidental generalizations of the considered judgments. The background theories themselves need to be more than reformulations of our considered moral judgments; indeed, those theories "should have a scope reaching beyond the range of the considered moral judgments used to 'test' the moral principles" (JP, 259). To be specific, within the Rawlsian framework, we have what Daniels labels 'level III' background theories, i.e., a theory of the person, a theory of procedural justice, general social theory, a theory of the role of morality in society, all of which persuade us to accept the contract apparatus and its constraints (which, in turn, enable us to select between competing conceptions of justice). Since these background theories are not a function of considered moral judgments about rights and entitlement, but appeal to other moral notions, such as claims about persons, these "level III theories provide a foundation for our notions of rights and entitlements without themselves appealing to such notions" (JP, 261).

This is a clever move by Daniels, a move that works only if it is true that there is no appeal, however tacit, to considered moral judgments concerning rights and entitlements. On this view, Rawls must be able to secure a foundation for our notion of rights and entitlements by appealing to claims about persons, procedural justice, social theory, and the role of morality in society. Such claims must be necessary and sufficient in order for Rawls to get a secure basis for the notion of rights and entitlements. The question as to whether or not Rawls succeeds in that attempt is outside the scope of this paper; what is important to note is that if this model is correct it would add justificatory force to those principles (avoiding circulatory and supplying the sorely needed independent criteria upon which to base both the principles and the contract model out of which they are chosen). But a dogged sceptic might well ask that, insofar as the level III theories are based on some considered moral judgments, why should we accept them? Even Daniels states that:

the level II apparatus (moral principles) will not be acceptable if competing theories of the person or of the role of morality in society are preferable to the theories Rawls advances. Rawls' Archimedean point is fixed only against the acceptability of particular level III theo-

ries . . . If the source of our disagreement about competing moral conceptions is disagreement on such level III considered judgments, then it is not clear just how much increase in tractability will result. (JP, 261, 264)

In responding to this problem and the subsequent cries of "subjectivism," Daniels takes pains to distinguish Rawls' system from that of the intuitionist theories that are foundationalist. Foundationalism carries with it the stigma of justifying moral principles by systematizing intuitions; such systematizing accounts for its apparent epistemic priority. Rawls' account, on the other hand, accords no such epistemic priority to considered moral judgments.

In addressing the complaint that reflective equilibrium merely systematizes some relatively determinate set of moral judgments, Daniels points out that wide reflective equilibrium permits extensive revision of these moral judgments. Even the judgment that "it is wrong to inflict pain gratuitously on another person" is revisable. "To imagine revising such a provisional fixed point, we must imagine a vastly altered wide reflective equilibrium that nevertheless is much more acceptable than our own. For example, we might have to imagine persons quite unlike the persons we know" (JP, 267). There is no set of judgments that is "fixed;" wide reflective equilibrium, as Daniels describes it, allows for drastic theory-based revisions of moral judgments. Not only are considered judgments about what is right and wrong revisable in light of background theories that conflict with them, so are those considered judgments that play a role in determining the acceptability of component level III theory.

Here is where Daniels runs into problems. He suggests we might revise the considered moral judgments that play a key role in forming our background theories if, for instance, "feasibility testing of the background theory" leads us to reject the theory and "therefore revise the considered judgment" (JP, 256). What exactly "feasibility testing" is he does not explain. He then suggests we might revise the considered judgment if the judgment is a part of our background theory "that is rendered implausible because of its failure to cohere with other, more plausible background theories" (JP, 256). But then we have our problem of coherence without necessarily having corresponding truth.¹² Third, he suggests that we would probably revise the level III judgment when it is part of a system of background theories that would lead us to accept principles we cannot reasonably accept. This is obviously problematic. First it is not clear we would (or ought to) give up the level III judgment; and more importantly, is the problem of determining the role of the considered judgment that gives rise to the back-

ground theories of the person. If we must, as Daniels contends, give up the judgment if it leads to a background theory that serves as the basis of principles we find unacceptable, a question that can be raised is: what are the criteria by which we determine whether a principle is acceptable? One would think the criteria would be the background theories and how the principles "fit" those theories. Which, in turn, leads to the question: why rely on those background theories? We can't rely on them because of their foundation, namely, considered judgments, since the system would then be open to attacks made on reflective equilibrium. Nor can we rely on them because they give rise to principles we find acceptable, for then we would have mere coherence.

Oblivious to these problems, or perhaps in spite of them, Daniels' main contention is that by distinguishing narrow reflective equilibrium from wide reflective equilibrium, Rawls escapes the charge of intuitionism and subjectivism. Daniels has come a long way in assisting the Rawlsian architectonic to defuse those charges, only to have his defense falter at a critical point. Without an adequate response to the question raised, it seems that Daniels, although instrumental in clarifying the conflict, fails in attempting to resolve it. It is not enough to claim that the initial judgments are constantly subjected to exhaustive review and are carefully tested. Unless we know precisely what it is against which those judgments are tested, we can be left with an aesthetically coherent, yet possibly false and insufficiently supported set of principles.

VI

Michael Sandel's main point in his Liberalism and the Limits of Justice, that an intersubjective dimension of the self is necessary for us to be the type of subject called for in Rawls' TJ, shifts our attention to the subject in the Rawlsian framework." Sandel claims that Rawls "require(s) a conception capable of marking out a wider subject of possession in which the subject is empowered to participate in the constitution of its identity" (LLJ, 152). This move is important because getting clear about the nature of the subject is absolutely crucial if we are to be in any position to make claims as to how a just social system is to be framed. Before we can begin to present a conception of justice within which principles are to be articulated, assign basic rights and duties, and determine the division of social benefits, we must be able to articulate a clear conception of the person for whom this conception is designed.

This brings us back to Daniels' paper and the unanswered question posed earlier. Why rely on background theories? We saw that we should not rely on them by

virtue of their originating through considered judgments, since that would raise the problems that concern Singer, Hare, and Brandt with regard to reflective equilibrium. Nor can we rely on them because they give rise to principles we find acceptable, since that would give us mere coherence, another problematic feature so readily vulnerable to attack.

Daniels' answer deserves a closer look. He maintains that wide reflective equilibrium may help us explain extensive agreement in our considered moral judgments, which is necessary for agreement on level III background theories. Thus far, this explanation is problematic for reasons stated earlier: agreement in our considered moral judgments does not necessarily involve any claims of truth concerning the background theories. But Daniels claims:

What wide reflective equilibrium shows us about the structure of moral theories may help us explain the extensive agreement we do find. Such agreement on judgments may reflect an underlying agreement on features of the component/background theories. Indeed, people may be more in agreement about the nature of persons, the role of morality in society, and so on, than is often assumed . . . (it may be) that the agreement is found because some of the background theories are, roughly speaking, true--at least with regard to certain important features moral agreement may not be just the result of historical accident, at least not in the way that moral disagreements are. Consequently, it would be shortsighted to deny credibility to considered judgments just because there is widespread disagreement on many of them; there is also agreement on many. Here moral anthropology is relevant to answering questions in moral theory. (JP, 272-73, my emphasis)

Clearly, we need to focus our attention on the status of our descriptive intuitions (what makes them reasonable or unreasonable, strong or weak). Through Boyd's analogical reasoning based upon scientific methodology, moral intuitions can be treated on the model of theory-determined intuitions in science. On that model, to the extent that background theories can be shown to be true, or approximately true, our intuitions that accord with those theories are reliable. After all, as Rawls recognizes, "(a)n intuitionist conception of justice is . . . but half a conception" (TJ, 41). We need ways, then, to support claims of truth, or approximate truth, regarding background theories.

Throughout the Dewey Lectures, Rawls wants to deny that a constructivist doctrine can approximate moral

truth. The parties in the original position are simply to select the conception most rational for them.

The parties in the original position do not agree on what the moral facts are, as if there already were such facts. It is not that, being situated impartially, they have a clear and undistorted view of a prior and independent moral order. Rather (for constructivism), there is no such order, and therefore no such facts apart from the procedure of construction as a whole; the facts are identified by the principles that result.¹⁴

But, of course, there are moral facts, if by moral facts we include the constraints necessary for one to be situated impartially in order to construct principles of justice that embody a prior conception of the person which underwrites the entire procedure. In other words, we deem the constraints imposed in the original position not as external, but as internally related to who we are qua moral beings. It might be true that we do not construct principles of justice that are moral facts or are expressions of judgments that are moral facts. But if we take the notion of pure procedural justice seriously, it would seem that we ought to arrive at principles that are moral in some sense (and not merely 'fair', as Rawls states). That is, they are constructed intimately from agents that have, embedded in their hypothetical existence, a moral nature representative of our moral nature.

Surely, we might agree with Rawls that by being situated impartially, the parties in the original position do not thereby have a clear and undistorted view of a prior and independent moral order. But Rawls does grant that the facts are identified by the principles that result; those principles are the outcome of a procedure that is based upon constraints viewed as internal to our moral nature. If Sandel is correct, and I believe he is, the independent standard "by which the reasonableness of our descriptive assumptions is assessed is given not by the laws of empirical psychology or sociology but instead by the nature of the moral subject as we understand it, which is to say that constitutive understanding we have of ourselves" (LLJ, 48).

It is this constitutive understanding we have of ourselves that serves as the foundation of the Rawlsian enterprise. Even Rawls says that justification of a conception of justice stems from that conception's "congruence with our deeper understanding of ourselves . . ." (KC, 519). We need only ask who are these agents in the original position. They are 'anyone'; 'I' am anyone in the original position insofar as 'I' have the essential characteristics of a human being qua moral agent and have, therefore, the wherewithal to choose

a conception that will hold for everyone once the veil is lifted and the characters of this world assume normal lives.

Sandel takes this task to be one of working backwards from the provisionally given principles of justice to the nature of the moral subject. Rawls focuses on the nature of the moral subject and argues through the original position to the principles of justice. Rawls feels we can take such liberty, since

the conditions in the original position are the fundamental principles governing our moral powers, or more specifically, our sense of justice; moral persons are entitled to equal justice . . . they are assumed to have a conception of their good . . . and are assumed to acquire a sense of justice We use the characterization of the persons in the original position to single out the kind of beings to whom the principles chosen apply . . . the capacity of moral personality is a sufficient condition for being entitled to equal justice . . . the principles that best conform to our nature as free and equal rational beings themselves establish our accountability . . . humankind has a moral nature (TJ, 51, 505, 519, 580).

The question is whether those who disagree with Rawls would do so on this level. As he notes, justification proceeds from what all parties to the discussion hold in common. Is it true that the conditions embodied in the description of the original position "are ones that we do in fact accept?" If we don't accept them, is it true that we "can be persuaded to do so by philosophical reflection?" (TJ, 21).

Peter Singer, Richard Brandt, and R. M. Hare disagree with Rawls, but do not ask these questions. Singer focuses on reflective equilibrium and claims that one need only describe one's moral capacity. Brandt argues that Rawls' system is merely a coherentist view of ethics. Hare claims that Rawls is a subjectivist insofar as the theoretical structure of justice as fairness is tailored to fit Rawls' intuitions and all that Rawls' system requires is that we, the readers, agree with those intuitions and, hence, begin with consensus.¹⁸ If Singer, Brandt, and Hare are to succeed in their attacks on the Rawlsian scheme, they must base their arguments on concerns affecting Rawls' notion of the subject, for it is from this descriptive vantage point that prescriptions flow.

Rawls anticipated the problems his critics have discussed: he notes that although a situation is one of stable equilibrium, it "does not entail that it is right or just . . . The moral assessment of each situation depends upon the background circumstance which

determine them" (TJ, 121). Further, Rawls does not doubt that there is an initial list of possible theories, expanded well beyond the list he considers, that would contain a better theory of justice than his own Two Principles. Nevertheless, to justify the system, we only need to show "that there is one interpretation of the initial situation which best expresses the conditions that are widely thought reasonable to impose on the choice of principles yet which, at the same time, leads to a conception that characterizes our considered judgments in reflective equilibrium" (TJ, 121). For Rawls, that very reasonableness stems from a description in the original position of the relevant background theories which is an accurate reflection of human moral circumstance. Our success in expressing our nature is a function of how consistently we act from our sense of justice. Our sense of justice "reveals what the person is, and to compromise it is not to achieve for the self free reign but to give way to the contingencies and accidents of the world" (TJ, 574-5).

Moral objectivity, in Kantian constructivism, "is to be understood in terms of a suitably constructed social point of view that all can adopt" (KC, 519). The key is that we have a hypothetical individual in the original position who is representative of all individuals, since such an individual fully captures the essential moral nature of persons. Singer, Hare, and Brandt might disagree with the restrictions placed in the original position: they might want to add, delete, or change some. But that is not their concern. To make the charge of subjectivism more convincing, however, our reflections have shown that it would be prudent to examine the background conditions that rest at the very center of the Rawlsian mode.¹⁶

NOTES

¹John Rawls, A Theory of Justice (Cambridge: Harvard University Press, 1971). Hereafter reference to this work will be abbreviated in the body of this paper, as TJ.

²John Rawls, "Outline of a Decision Procedure for Ethics," Philosophical Review 60, 1951, pp. 177-197.

³Peter Singer, "Sidgwick and Reflective Equilibrium," The Monist 1974, p. 516.

⁴R. M. Hare, "Rawls' Theory of Justice," in Reading Rawls, edited by Norman Daniels. (New York: Basic

Books, Inc.), p. 82. Hereafter designated in the text as RTJ.

⁹Singer, "Sidgwick and Reflective Equilibrium," p. 494. Hereafter designated in the text as SRE.

¹⁰Questions concerning these background conditions will prove to be of utmost concern when we look at Norman Daniels and others of his ilk.

¹¹Richard Brandt, A Theory of the Good and the Right (Oxford: Clarendon Press, 1977. Hereafter designated as TGR.

¹²As Larry Thomas has pointed out in his comments on an earlier draft of this paper, this is a rather problematic claim. Rawls recognizes that "even in a well-ordered society there are some persons for whom the affirmation of their sense of justice is not a good" (TJ 575).

¹³Rawls claims that the principles chosen in the original position are objective in that they "are the principles that we would want everyone (including ourselves) to follow were we to take up together the appropriate general point of view. The original position defines this perspective, and its conditions also embody those of objectivity: its stipulations express the restrictions on arguments that force us to consider the choice of principles unencumbered by the singularities of the circumstances in which we find ourselves (We) take up a point of view that everyone can adopt on an equal footing. In this sense we look at our society and our place in it objectively: we share a common standpoint along with others and do not make our judgments from a personal slant. Thus our moral principles and convictions are objective to the extent that they have been arrived at and tested by assuming this general standpoint and by assessing the arguments for them by the restrictions expressed by the conception of the original position" (TJ 516-17).

¹⁴Ronald Dworkin, "The Original Position," in Reading Rawls, edited by Norman Daniels. (New York: Basic Books, Inc.), pp. 16-52. Hereafter designated as OP.

¹⁵Norman Daniels, "Wide Reflective Equilibrium and Theory Acceptance in Ethics," Journal of Philosophy, 1979, p. 257. Hereafter designated as JP.

¹⁶This is not addressed by Daniels here, although he does have something to say about this in a different context.

¹³Michael Sandel, Liberalism and the Limits of Justice (Cambridge: Cambridge University Press, 1982). Hereafter designated as LLJ.

¹⁴John Rawls, "Kantian Constructivism in Moral Theory," Journal of Philosophy 77, 9, 1980, p. 568.

¹⁵Rawls writes: "(O)ur object should be to formulate a conception of justice which, however much it may call upon intuition, ethical or prudential, tends to make our considered judgments of justice converge. If such a conception does exist, then, from the standpoint of the original position, there would be strong reasons for accepting it, since it is rational to introduce further coherence into our common convictions of justice" (TJ 45). I would maintain that insofar as we desire only that considered judgments of justice converge, without presupposing any reasons for such convergence, the claims that the Rawlsian system collapses into merely a coherentist view are strengthened. From this passage cited above, it seems that we have strong reasons for accepting a conception of justice so long as our considered judgments of justice converge in the OP. Such an interpretation, although one which Rawls seems to allow in such passages is not a fair representation given all of the constraints in the original position.

¹⁶I'd like to acknowledge the many helpful conversations with Maynard Adams, Steven Darwall, and Larry Thomas without which this essay would be more deeply flawed than it is presently.