
Rawls's Justification Model for Ethics: What Exactly Does It Justify?

Necip Fikri Alican

Rawls is famous for two things: his attempt to ground morality in rationality and his conception of justice as fairness. His work has been resounding on both fronts, the first constituting the justificatory framework for the second. Yet from the beginning, the outcome has been more doctrinaire than the method should have allowed with design details promising objectivity. This article goes to that beginning, or to a reasonable proxy for it, in the "Outline of a Decision Procedure for Ethics," with the aim of exposing and examining the discrepancy where it originates. The goal is not to prey on the earliest version of an initiative later undergoing revision but to identify and investigate the inception of a systematic bias that is retained rather than revised in subsequent iterations. The primary finding is that justice as fairness is too dogmatic an outcome for the decision procedure proposed. The corresponding principles (of justice as fairness) are consistent with the proposed procedure only relative to a specific system already presumed valid. A secondary finding, supported by the same considerations as the first, is that the decision procedure itself, regardless of whether it works well with the particular principles in question, may not be as use-

NECIP FIKRI ALICAN is the author of *Mill's Principle of Utility: A Defense of John Stuart Mill's Notorious Proof* (Atlanta: Brill, 1994) and *Rethinking Plato: A Cartesian Quest for the Real Plato* (New York: Brill, 2012).

ful as it appears, turning out to justify only, or mostly, what is widely accepted anyway.

1. *Razzle-Dazzle*

We all like a good story. And if it is told just right, it does not even have to be true. The crystal ball, the deck of cards, the coffee grounds—they are all extraneous to the fortune told. The divination itself does not come from them. The oracle may, as it happens, be right on target, but it is the show that promotes persuasion. Without the razzle-dazzle, what the diviner has is just another story, true or false, right or wrong, good or bad. The purpose of this article is to determine where Rawls gets his story, how credible it is, and whether it works.

This is not an assault on Rawls. He tells a good story. But it does recall the allegorical introduction above. Essay after essay, book after book, everything comes out just right, that is, his own moral outlook turns out to be validated by an ethical justification model of his own design. There would be nothing strange in this if it were not for Rawls's assurance that the model guarantees objectivity. As it is, given the claim to objectivity, what is suspicious is that an objective approach should be so biased toward a particular outcome, or if one prefers, toward a particular kind of outcome.

A tempting response is that a good model would naturally be biased toward the right answer. Another is that ignoring or weeding out wrong answers is not a violation of objectivity. There is indeed nothing wrong with giving the same answer every time if it is the right answer. But that is precisely what is in question.

To restate the problem, this time in more formal terms than in the opening allegory, Rawls operates with a decision procedure for ethics that keeps corroborating the same moral outlook, a liberal one, whereas the objectivity he claims for the procedure might reasonably have been expected to be consistent with a wider range of moral, social, or political perspectives, or perhaps with a single position equidistant to polar extremes.¹ The aim of this article is to ascertain whether this

*Rawls's
justification
model
promises
objectivity
but keeps
delivering
liberalism.*

¹ I am merely acknowledging, as opposed to either praising or protesting, the liberalism of Rawls, who openly embraces both the position and the name for it. The label "liberal," much like the label "conservative" with which it is

is a real problem and whether it may be overcome to salvage the analytic apparatus Rawls recommends for ethical justification.

Departure from objectivity is, of course, always a real problem. And this is the conclusion of this article as well. But its focus is more specific. It is concerned with two related questions. The first is about internal consistency: (1) Does the ethical justification model developed by Rawls justify the moral principles advocated by Rawls? The second is about universal relevance: (2) Can the ethical justification model developed by Rawls justify any moral principles?

Covering both questions, the overarching reason for suspicion, to be confirmed or rejected here, is that certain methodological constraints, conceived as anchors to objectivity (freedom from bias) and safeguards against relativism, may not just undermine the justification of the particular principles presented by Rawls, thus making him dogmatic in endorsing them, but also preclude the justification of any moral principles whatsoever, or at least any useful or substantive ones (the kind directly relevant to the adjudication of moral problems). In the first case, the orientation toward objectivity and the precautions against relativism contradict the principles adopted. In the second case, they engender sterility: the justification of generalizations that are already widely accepted.

This is not, even incidentally, a rejection of the principles espoused by Rawls. More accurately, it is not a critique of those principles independently of their relation to the justification model employed to validate them. The principles themselves are neither surpassed by alternatives nor in urgent need of justification. I myself would gladly defend them on that point, that is, from a standalone perspective separated from the question of justification by Rawls, but I will not do so, not here anyway, as this is an altogether different matter. I will, however, question their justification through the decision procedure developed and proposed by Rawls.

Let me acknowledge at once that both the model and the

regularly contrasted in ordinary discourse, can become charged with negative connotations and thereby associated with negative intentions. As against this, the reference here is strictly in recognition of the liberalism Rawls claims for himself: 1985, 245–251; 1987, 1–25, especially 18, including n. 27; 1993, 6, 156–158, 233; 2001, 104–106, 148–150, 153–157, 189–192.

output have evolved over time. I do not deny this. What I deny is that the evolution has made my concerns irrelevant. Granted, ethical justification has always been a work in progress for Rawls. But my impression is that the problems identified in this article were never met at any point in that process because they were never even on the agenda. This impression is not a hunch but an observation, or to be precise, a judgment formulated on the basis of a series of observations. While negative claims are notoriously difficult to prove, section 5 of the present article is a representative effort. The aim there is not to establish beyond a doubt that Rawls later says nothing that could possibly be interpreted as an improvement upon any of the difficulties in his earlier works but to demonstrate a relevant methodological continuity between the earlier and later works that corroborates the timeless relevance of my concerns here.

Evolution of Rawls's model and output over time has not rendered present concerns irrelevant.

To be sure, the model's evolution can be traced through various paradigms in succession: the considered moral judgments of competent moral judges; the strategic outlook of contractors working out a reflective equilibrium from the original position; the various modes of public reason (or public justification) converging toward an overlapping consensus (grounded in a reasonable moral psychology). The corresponding normative output undergoes development as well, ranging from the relatively detailed set of principles presented early on to the more elegant and now famous twin principles explicating Rawls's conception of justice in social institutions.

Regular recalibration on the part of Rawls requires meticulous attention on the part of the critic. The broader the scope of consideration, the better. But that is a project for a book, at least one, probably more. A feasible alternative to comprehensive coverage is to work with *A Theory of Justice* (1971/1999, hereafter "*Theory*").² That might be the obvious choice, but that work has undergone enough critical scrutiny to save generations of scholars the trouble of looking for something new in it. Besides, it remains in the middle, drawing on earlier material and anticipating later development, but representing

² Parenthetical references giving a publication year (and, where relevant, page numbers) without naming the author are to Rawls. Quotations omit footnote reference markers in the original.

neither perspective adequately. Fresher insight can be had by going back to the beginning, which cannot be too far from Rawls's "Outline of a Decision Procedure for Ethics" (1951, hereafter "Outline").³

Starting from the beginning should prove enlightening (section 2) so long as observations and inferences (sections 3–4) are filtered through an awareness of subsequent departures and turning points (section 5). Proceeding on that basis, this article examines, first whether Rawls's decision procedure for ethics justifies the moral principles he purports to be justified by that model, second whether that procedure facilitates the adjudication of actual moral problems (perennial or emerging) or remains an exercise in formalism with little or no practical import or serviceable implications.

The conclusion regarding the first question is that the validation connection between the model and the output is not as strong, or even as clear, as Rawls would have us believe, and that Rawls is, in this respect, inevitably dogmatic in his support of the principles he advocates. The conclusion regarding the second question is that design elements intended to ensure objectivity and thus to avoid relativism weaken the credibility of the model as a decision procedure relevant to moral reality, or to be more specific, as a decision procedure capable of justifying the kind of moral principles that would be useful in the adjudication of moral problems.

This, in effect, is to leave Rawls with a vacuous decision procedure and dogmatic principles. And that, in turn, may seem to underestimate his contributions. But it is not as harsh as all that. While there is indeed some tension between this criticism and Rawls's reputation, there is also room enough for both. Nothing is ever gained from reacting to mediocrity. And no harm can come from exposing a problem in excellence.

³ While the "Outline" (1951) is indeed Rawls's first publication, it is also common to start with his doctoral dissertation (1950), and sometimes even with his undergraduate thesis (1942). This is just as well, since the "Outline" originates in his dissertation. As for his senior thesis, that piece was bundled together with a short manuscript from Rawls's maturity and published posthumously as a book (2009), with notable philosophers serving as editors and commentators, including Habermas with an afterword for the German edition (2010). His dissertation has had a steadier audience, but it, too, has recently seen something of a revival in readership and commentary. See, for example, Mäkinen and Kakkuri-Knuuttila (2013).

Criticizing Rawls detracts no more from his contributions than centuries of criticism have from Plato's.

2. *Immediate Context*

Here is the immediate problem motivating Rawls to seek a solution later becoming the focus of his life's work:

Does there exist a reasonable decision procedure which is sufficiently strong, at least in some cases, to determine the manner in which competing interests should be adjudicated, and, in instances of conflict, one interest given preference over another; and, further, can the existence of this procedure, as well as its reasonableness, be established by rational methods of inquiry? [1951, 177]

This particular quotation is from the "Outline" but the question itself has been with Rawls throughout his career. He approaches the problem in two parts. First (1951, 177–191), in the theoretical portion of his outline, he develops a general decision procedure for the justification of moral principles. Second (1951, 191–197), in the practical portion of his outline, he uses that decision procedure to justify a specific set of moral principles he introduces as "principles of justice."⁴ With respect to the first part, his decision procedure states that moral principles are justified insofar as they explicate moral judgments that coincide with the considered moral judgments of competent moral judges (1951, 187–189).⁵ With respect to

Moral principles are justified insofar as they explicate moral judgments that coincide with considered judgments of competent moral judges.

⁴ These are the seven moral principles in section 5.5 of the "Outline" (1951, 192–193). They mark the inception of Rawls's conception of justice as fairness. This is a dynamic conception, evolving throughout his career, but not so much that it cannot be identified (wherever it turns up) as essentially the same general notion, and characterized as a conception of justice as fairness, in any formulation and combination of principles used to express it: 1951, 192–193; 1957, 653–654; 1958, 165; 1971, 60, 302, cf. 83, 250 (=1999, 53, 266–267, cf. 72, 220); 1982, 5, cf. 46–55; 1985, 227; 1993, 291, cf. 331–334; 2001, 42. The most prominent turning point in the course of evolution is the transition from seven (1951) to two (1957/1958) principles. But that is not the end of the story. The twin principles themselves change over time. Note, for example, that Rawls identifies (1993, 5, n. 3) the Tanner lectures (1981/1982) as a corrective transition between the principles of justice in *Theory* (1971, 60, 302 [=1999, 53, 266–267]) and those in later works (1982, 5, 46–55)—a correction largely in reaction to criticism from Hart (1973). Nevertheless, this does not rule out the continuity of the principles as increasingly more satisfactory attempts (at least as far as Rawls is concerned) to investigate the same concept.

⁵ Rawls explains in detail what he means by "competent moral judge" (1951, 178–181), "considered moral judgment" (1951, 181–183), and "explication"

the second part, his moral principles explicate a conception of justice as fairness in social institutions—or they explicate moral judgments that hang together as a conception of justice as fairness in social institutions (1951, 192–193).⁶

This section of the present article is devoted to an assessment of the compatibility between the two parts of Rawls's paper in an effort to determine both whether his decision procedure validates his conception of justice and whether it is capable of justifying anything at all that might prove normatively useful as part of a comprehensive moral outlook. The reason for hesitation, as explained in general terms in the preceding section, is that the conception of justice is a liberal one, whereas the decision procedure is one that should, given the emphasis on objectivity, yield principles that have a more universal appeal.⁷ Moral principles making up a liberal conception of justice are unlikely to be filtered without loss

(1951, 184–187). These have long been standard terminology in the literature. However, apart from these, the summary statement encapsulating the decision procedure, the one to which this note refers, is rough in the sense that it includes one and omits three of what Rawls cites as jointly sufficient criteria for the justification of moral principles. The other three criteria (1951, 187–189) are as follows: (1) A moral principle must show a capacity to become accepted by competent moral judges after they have freely weighed its merits by criticism and open discussion and after each has thought it over and compared it with his or her own considered moral judgments. (2) A moral principle must be able to function in existing instances of conflicting opinion and in new cases causing difficulty. (3) A moral principle must show a capacity to hold its own against a subclass of the considered moral judgments of competent moral judges. The summary statement still suffices for the fundamental notion behind Rawls's account of the justification of moral principles. The other three criteria are best brought out as they become relevant to the discussion.

⁶ While “justice as fairness” (1957, 1958, 1985, 2001) is not yet standard terminology in the “Outline,” the moral principles in that work are, by Rawls's own designation, “principles of justice” (1951, 191). And he is already concerned there with “fair decisions on moral issues” (1951, 181) and with a “fair opportunity for all concerned” (1951, 182). Hence, even without a formal name for them, the seven principles in question are, in fact, principles of justice as fairness (cf. n. 4 above).

⁷ As stated in n. 1 above, the term “liberal” is intended in a descriptive sense faithful to the actual liberalism Rawls champions in print. The “Outline” predates the explicit advocacy of liberalism present in the references cited above (n. 1), but the orientation is the same, implicit at least in the continuity between the seven principles in the “Outline” and the twin principles formulated later (cf. nn. 4, 6 above). The evidence for this is the conception and presentation of “justice as fairness” (and thereby any formulation, early or late, of the attendant principles) as “a form of political liberalism” (e.g., 2001, 183).

through a decision procedure that is as impartial as claimed. This specific source of apprehension recalls the more general one of whether such filtration would leave us with anything worthwhile at all.

Not all competent moral judges across time, nations, and cultures can reasonably be expected to have an identical set of moral judgments.⁸ Not even all competent moral judges within the same time period, nation, and culture are likely to have an identical set of moral judgments. This suggests that a decision procedure drawing on the wisdom of a panel of judges needs certain safeguards to keep relativism from creeping into the system. Rawls's precaution against relativism consists of universalistic clauses built into the decision procedure. Design specifications call for both considered moral judgments and reasonable moral principles that transcend time, place, and person.

Concerning considered moral judgments, Rawls requires, among other things,⁹ a demonstrable relevance and effectiveness inherent in the judgment as a stability of sorts:

It is required that the judgment be stable, that is, that there be evidence that at other times and at other places competent judges have rendered the same judgment on similar cases, understanding similar cases to be those in which the relevant facts and the competing interests are similar. The stability must hold, by and large, over the class of competent judges and over their judgments at different times. Thus, if on similar cases of a certain type, competent judges decided one way one day, and another the next, or if a third of them decided one way, a third the opposite way, while the remaining third said they did not know how to decide the cases, then none of these judgments would be stable judgments, and therefore none would be considered judgments. These restrictions are justified on the grounds that it seems unreasonable to have any confidence that a judgment is correct if competent persons disagree about it. [1951, 182–183]

⁸ Rawls's later works promote pluralism, but the result is still (in Rawls's own words) a "liberal conception of justice" (e.g., 1987, 18–22; 1993, 144–168). The pluralism espoused in 2001 (3–5), for example, employs an "overlapping consensus" (2001, 32–38, 192–195; cf. nn. 19, 20 below) coinciding with the general outlook supported by the moral principles advocated fifty years earlier (1951, 192–193). This is a common theme in the later works, where pluralism invariably comes with "liberalism" as a shared value or emergent mindset. This is taken up in greater detail in section 5 of this article.

⁹ The reference in "among other things" is to the other six of the seven criteria Rawls presents for considered moral judgments (1951, 181–183).

*Stability
requirement
for considered
moral
judgments.*

This is Rawls's sixth criterion for considered moral judgments. Call it, to give it a name, the "stability requirement" (for considered moral judgments). This is not just for the sake of convenience in reference but also in conformity with his own usage as seen in the first sentence of the passage quoted above.

Being sensitive to this criterion requires special terminology to avoid self-contradiction in discussing the decision procedure. For example, it would be contextually meaningless, as well as factually wrong, to claim that Rawls ignores the possibility that the considered moral judgments of competent moral judges may vary across time and persons. This is because considered moral judgments are not the kinds of things that can vary in this way. They are, by definition, the kinds of moral judgments on which there is unanimous (or near-unanimous) agreement. Thankfully, avoiding semantic obstacles to what might otherwise turn out to be a fruitful discussion is a simple matter of making room for a new term, say, "sensible moral judgment," referring to those moral judgments that satisfy all of Rawls's criteria for considered moral judgments except the stability requirement. A considered moral judgment, then, is a sensible moral judgment on which all or nearly all competent moral judges agree.

Concerning reasonable moral principles, Rawls likewise requires a broad and enduring appeal in addition to the satisfaction of various other conditions:¹⁰

Thirdly, the reasonableness of a principle is tested by seeing whether it can function in existing instances of conflicting opinion, and in new cases causing difficulty, to yield a result which, after criticism and discussion, seems to be acceptable to all, or nearly all, competent judges, and to conform to their intuitive notion of a reasonable decision. . . . In general, a principle evidences its reasonableness by being able to resolve moral perplexities which existed at the time of its formulation and which will exist in the future. [1951, 188]

This is Rawls's third criterion for reasonable moral principles. This, too, is a kind of stability requirement. It looks for a stability of sorts in moral principles. Reasonable moral principles should be useful both under existing conditions

¹⁰ The other conditions invoked here are the remaining three of a total of four criteria Rawls requires of reasonable moral principles (1951, 187–189).

and under other circumstances, at present or in the future. A duplication of terms, using the same word to refer to two different criteria, can lead to confusion. Instead of reusing “stability,” this criterion might better be called, again for the sake of convenience in reference, the “reliability requirement” (for reasonable moral principles).

Reliability requirement for reasonable moral principles.

The two requirements presumably shield the system from relativism. The stability requirement for considered moral judgments seems to be designed for the express purpose of preventing relativism from creeping into the decision procedure. The reliability requirement for reasonable moral principles seems to be designed to prevent ad hoc justifications of moral principles. The two requirements thus appear to be conceived as severally necessary, and perhaps jointly sufficient, to avoid relativism.

However, the clauses of universalism designed to ward off relativism threaten the compatibility of theory and application in Rawls’s outline and diminish the usefulness of the decision procedure itself. If all or nearly all competent moral judges at all times in all places were to agree on a set of sensible moral judgments, that might indeed rule out relativism. But would these competent judges actually endorse Rawls’s principles of justice? Would they even agree on anything at all, or at least on anything morally substantive? The questions point to two problems.

First, Rawls’s decision procedure is not likely to justify his own moral principles, because these particular moral principles, which constitute a conception of justice as fairness, do not seem to explicate moral judgments that coincide with the sensible moral judgments of all or nearly all competent moral judges at all times in all places. The decision procedure requires competent moral judges to be in agreement, but not all, or even nearly all, competent moral judges will endorse the liberal conception of justice Rawls lays out as he demonstrates where the theory takes us in practice. If not, then Rawls’s particular moral principles cannot reasonably be said to explicate sensible moral judgments that meet the stability requirement. The question here is whether and how Rawls can deny, avoid, or justify a kind of dogmatism in advocating moral principles that explicate only some of the sensible moral judgments of

Rawls’s decision procedure unlikely to justify his own moral principles.

only some of the competent moral judges. The term “dogmatism” is intended not in any special sense, philosophical or otherwise, but in the ordinary sense of adopting a rigid position without sufficient evidence, or against evidence to the contrary, for example, in this case, including the likelihood that different competent moral judges would have different sensible moral judgments.

*Danger of
vacuity.*

Second, it is possible that the clauses of universalism in Rawls’s decision procedure are so strong that the method he outlines is vacuous. It is possible, in other words, that no moral principles (and certainly no normatively useful ones) explicate moral judgments that coincide with the sensible moral judgments of all or nearly all competent moral judges at all times in all places. This is a more fundamental problem than the first because it underscores the possibility that, regardless of whether Rawls’s moral principles are good ones, or ones justified by his own decision procedure, the decision procedure itself may not be very useful. Note that it is not necessary to consider this problem in its strongest form to appreciate its gravity. It would be just as problematic for Rawls’s decision procedure if very few moral principles, or very few useful, significant, and substantive ones, were to explicate moral judgments that coincide with the sensible moral judgments of all or nearly all competent moral judges at all times in all places.

Can Rawls, faced with these two problems, preserve his decision procedure and his principles of justice in the form in which he presents them?¹¹ The remainder of this article is

¹¹ A third problem, actually a different way of thinking about the first problem, is the scenario where the decision procedure justifies Rawls’s moral principles but also at least one set of moral principles in conflict with those of Rawls. The assumption of unanimous (or near-unanimous) agreement makes this scenario irrelevant because unanimous or near-unanimous agreement on mutually inconsistent principles does not present a meaningful context for discussion. Anyone rejecting that assumption, however, could take up the first problem as two separate ones: (1) the possibility that the decision procedure does not justify Rawls’s moral principles; (2) the possibility that the decision procedure justifies Rawls’s moral principles along with alternatives to Rawls’s moral principles. There is, of course, no good reason to reject the prior assumption, since the appeal to competent judges can hardly be helpful without the assumption of substantial agreement, whether this be unanimous, nearly unanimous, widespread, or strong. Note also that, if the assumption

devoted to this question. To reiterate the general direction, the answer is that he might be able to preserve them to a degree, holding on to fundamentals, but that he must make concessions. Regarding the first problem, he must admit to substantial dogmatism and look outside the system for a way to justify the dogmatism as opposed to denying it or trying to avoid it while remaining within the system. Regarding the second problem, he must relax the universalism in his decision procedure to alleviate (but not eliminate) the difficulty that it can be vacuous and otiose as a method of ethical justification. He could perhaps reposition the decision procedure as a justification model for a community of sufficiently like-minded moral agents rather than a universal platform (the two being the same if everyone happens to be in that community).

Before establishing the existence and implications of dogmatism in the way Rawls promotes the principles he does in the "Outline," it may help to examine his claim to objectivity in the same work:

[T]he objectivity or the subjectivity of moral knowledge turns, not on the question [1] whether ideal value entities exist or [2] whether moral judgments are caused by emotions or [3] whether there is a variety of moral codes the world over, but simply on the question: [4] does there exist a reasonable method for validating and invalidating given or proposed moral rules and those decisions made on the basis of them? [1951, 177]

According to Rawls, then, what makes a method objective is that it is reasonable. But this is a technical conception of objectivity. It is not the kind of objectivity that precludes dogmatism, which need not be unreasonable, nor even unjustifiable. Might not conflicting principles be supported equally well by equally reasonable decision procedures, perhaps even by the very same one, thus making any choice dogmatic—or, worse, arbitrary—and in that sense and to that extent "non-objective"? Rawls himself admits that his decision procedure may not be the only reasonable one:

There is no way of knowing ahead of time how to find and

were rejected, this would remove the grounds originally invoked for suspecting the decision procedure might not justify Rawls's moral principles. Hence, the first problem would, in that case, be transformed not into the two separate problems described in this note but into just the second of those two problems.

formulate these reasonable principles. Indeed, we cannot even be certain that they exist, and it is well known that there are no mechanical methods of discovery. In what follows, however, a method will be described, and it remains for the reader to judge for himself to what extent it is, or can be, successful. [1951, 178]

The decision procedure Rawls outlines is, in fact, objective in the standard sense, that is, in the sense of impartiality (whether or not it is also objective in the sense depicted above). And its objectivity is so clear as to require no further argument. There is no inherent breach of impartiality in the position that moral principles are justified insofar as they explicate the considered moral judgments of competent moral judges. This is not to say that the decision procedure cannot be manipulated to favor certain outcomes or to suppress others. But that sort of departure from objectivity is grounded in the manipulation of the system and not in the system itself. The decision procedure is and remains objective.

Application is a different matter. Even if Rawls cannot reasonably be held responsible, as the designer of the system, for any breakdown or transgressions in practice, especially for any in the form of misuse, he could still be faulted, as an ordinary operator, for abusing or misusing the system during application. This is the possibility that Rawls the moral agent may be at fault even where Rawls the moral engineer is not. As it turns out, however, and as I aim to show, he needs to reconsider what he does in both roles.

The following two sections elucidate these two issues. The next section (section 3) is on the specific matter of the application bias keeping Rawls's distinctive moral principles from being justified by his universalistic decision procedure. The section after that (section 4) is on the general problem concerning the possible sterility of the decision procedure as a model of ethical justification.

3. Application Bias

This section demonstrates that Rawls's principles of justice are dogmatic in the sense that they are not justified by his decision procedure, which requires the unanimous (or near-unanimous) agreement of competent moral judges, and further, that

this is a real problem he can neither deny nor avoid without altering either his principles or his decision procedure or both. The reasonability of denial and the possibility of avoidance are worth considering separately and seriously as either one would absolve Rawls of the charges advanced here.

*Rawls's
principles of
justice not
justified by
his decision
procedure.*

Denying the dogmatism is an almost instinctive response in this case. This requires not just rejecting the charge but also giving good reasons for doing so. One way of pursuing this possibility is to start with a context of moral conflict where the principles espoused by Rawls clash with those espoused by others. The acid test here is whether those other principles have a comparable claim to reasonableness.

A setting of historical and cultural conflict presents suitable testing grounds. This is not because the problem itself is specifically historical or inherently cultural but because superimposing historical and cultural dimensions can help visualize the problem better at least as a first approximation. Suppose, then, that the task is to compare judges and judgments across history and between cultures. Consider, to begin, the historical dimension, which will almost always bring with it a cultural perspective: Do Rawls's moral principles explicate moral judgments that coincide with the sensible moral judgments of historically competent moral judges? They do not. His principles of justice, even in their earliest versions (1951, 192–193), and perhaps more clearly in their later versions in subsequent works, explicate moral judgments that take justice as fairness. Such an account stands in contrast to some older conceptions of justice, which involve the production and preservation of distinctions in class, race, and gender, and which were generally considered to depend for their justification, at least in part, on a conviction that they reflect the natural order of the world.

We may note, for example, that the ancient Greeks held moral principles that explicate sensible moral judgments allowing and even requiring class distinctions. The problem for Rawls would then be to justify dismissing the sensible moral judgments of competent moral judges in, say, Periclean Athens in favor of the moral judgments of their more liberal descendants in modern-day Western democracies.

The solution seems simple enough: The people of Periclean Athens were not competent moral judges and their moral

judgments are therefore irrelevant. This makes them neither incompetent nor immoral, pointing instead to a technical disqualification: They do not satisfy the criteria pertinent to the designation “competent moral judge” (1951, 178–181). Do they all fail in this regard? Probably not, but that is not important. The point is not that no one in Periclean Athens could possibly have been a competent moral judge but that Periclean Athens is not the place to go looking for competent moral judges.

This otherwise appealing answer is nevertheless unacceptable because it begs the question of the competence of the relevant judges. It is wrong to dismiss the sensible moral judgments of the people of Periclean Athens merely because their conception of justice or their moral outlook is repugnant to our own moral sensibilities. Doing so violates the rules of Rawls’s own decision procedure:

a competent judge has not been defined by what he says in particular cases, nor by what principles he expresses or adopts. . . . Obviously if a competent judge were defined as one who applies those principles, this reasoning would be circular. Thus a competent judge must not be defined in terms of what he says or by what principles he uses. [1951, 180]

The alternative is to refuse to recognize as a competent moral judge anyone who fails to satisfy all the criteria for competent moral judges. As for the judgments of those who fully qualify as competent moral judges, Rawls may further refuse to recognize as a sensible moral judgment any judgment that fails to satisfy all the criteria for sensible moral judgments.¹² The pertinent question in this case is whether all (or nearly all) judges and judgments that contradict the conception of justice as fairness fail to satisfy some criterion or other. A case of this sort may be made piecemeal against this or that particular challenge, but such a recalcitrant approach is not sustainable as a general defense.

Could Rawls perhaps invoke the stability requirement, the very criterion that marks the difference between “sensible”

¹² Strictly speaking, sensible moral judgments have no formal criteria, at least not in the original work. Rather, they are invented here, in the present article, for the sake of circumventing a specific semantic obstacle in the discussion. That said, the derivative criteria for sensible moral judgments can be taken as the formal criteria for considered moral judgments minus the stability requirement.

and “considered” moral judgments? Could he contend that Periclean moral judgments fail the stability test, as they clearly have no place in our world, where a liberal moral outlook and a conception of justice as fairness prevail? But whose world is that? Where exactly does this outlook, this conception, prevail? Yes, the old makes room for the new. And the ways of the ancient Greeks are a thing of the past. But much of what is problematic (for Rawls’s decision procedure) about Periclean Athens is readily instantiated in modern day India to take just one contemporary example. At the very least, the class distinctions are still with us. How is this not “our” world?

India is just one example, chosen because the culture there is one of the great traditions familiar everywhere. But comparable problems can be found elsewhere. People in various Middle Eastern cultures, for instance, still sell their daughters into marriage shortly after their first menstrual cycle, as arranged through prior verbal contract, typically upon the birth of the child. Justice demands it. Is there room for this kind of justice in Rawls’s decision procedure? There should be.¹³ But we never see it with Rawls in charge of application.

Later refinements in Rawls’s metaphors and thought experiments for ethical justification help obviate counterexamples such as modern day slavery, an institution unlikely to find much support behind a veil of ignorance. But counterexamples do not have to be focused tightly on class distinctions or on social practices or institutions bordering on slavery to be successful. It should not be too hard to come up with a scenario that divides competent moral judges (and even contractors behind a veil of ignorance as envisaged later) without so extensively contradicting our current sensibilities as the thought of slavery does.

One can surely imagine a moral, social, or political platform

¹³ I am not claiming that selling one’s daughter into marriage is just. Nor am I claiming that a good decision procedure for ethics should make room for selling one’s daughter into marriage. I am claiming that, given the exposition of Rawls (1951), his decision procedure for ethics should (by design) have room for selling one’s daughter into marriage if that is something agreeable to competent moral judges. The scenario should not be ruled out before it is processed through the decision procedure. Any rejection should instead be the result of such processing. The case should be taken to the judges instead of the judges being selected on the basis of how they would decide the case.

that is, unlike class distinctions or outright slavery, consistent with Western ideals yet inconsistent with Rawls's favored worldview. Or leaving Rawls out of the equation, there could conceivably be two or more competing (mutually inconsistent) conceptions, outlooks, or paradigms each of which is broadly consistent with sociocultural norms in modern Western societies.¹⁴ In political economy, for example, both the principles of social welfare and the principles of *laissez-faire* might explicate moral judgments that coincide with the sensible moral judgments of competent moral judges. Professional economists and lawmakers may (and actually do) disagree not only on policy issues but also, and more fundamentally, on the right choice among competing economic systems. Yet nothing in that choice signals a violation of any of the qualification criteria for Rawlsian judges or judgments.¹⁵

Rawls proceeds as if people of certain qualifications meditating in a certain way would never arrive at moral judgments that can be explicated by contradictory or competing moral principles. This is an axiomatic component of the decision procedure. But if the example from comparative economic systems is a good one—and if it is not, one that is may still be found—then the assumption is not justified. We may safely predict that a good many conflicts can be resolved by showing that the opposing view fails to satisfy some criterion or other in the Rawlsian framework. But this cannot be a blanket answer. It is unreasonable to insist that any and every conflict with Rawls's principles of justice is due to the opposing view's failing to meet a legitimate condition.

Can Rawls instead avoid the dogmatism before it happens, as opposed to denying it after it does, by reformulating his decision procedure to justify his principles while at the same time salvaging its universalistic character? One way to

¹⁴ For the sake of argument, let us set aside the debate on whether a typical modern Western society is a real phenomenon or a heuristic device to speak of norms and averages while instead dealing with a heterogeneous data set where nothing, or not much at all, actually corresponds to what is taken to be typical. I will gladly yield on this point.

¹⁵ Rawls would likely subordinate the principles of free enterprise to those of state intervention in social welfare, but he could not, if he were to make that choice, justify it through his decision procedure. There is nothing in that procedure either to facilitate or to validate that kind of choice.

attempt this might be to redefine the criteria for competent moral judges to bypass the problem of dogmatism, perhaps by stipulating that a competent moral judge be familiar with all or nearly all philosophies and philosophers, thereby taking all their sensible moral judgments into account while being biased neither toward nor against these judgments.¹⁶ The purpose of such an amendment would be to ensure as far as possible that every sensible moral judgment gets a fair shot at being admitted into the class of considered moral judgments. This, in turn, would presumably avoid dogmatism while preserving both the decision procedure and the principles of justice (as fairness).

But the level of competence required would be nothing short of sagacity. Some sort of limiting scope might help ease the burden. What principles of selection should be used to determine the areas of expertise to be required of a competent moral judge? Or is this really a question in the right direction? True, a competent moral judge is unlikely to know everything required to satisfy an amendment requiring an open-ended expansion of the base of competence and expertise. On the other hand, being selective about the scope of knowledge to be sought seems to perpetrate the dogmatism such an amendment would be designed to avoid in the first place. One way out of this dilemma is to make use of the plurality of judges. Rawls does not limit himself to a single “ideal observer.” He speaks of “competent moral judges” in the plural. Perhaps we can construe this as a panel of competent moral judges in which each judge is an expert in one area. In this way, it might be possible to maximize coverage of the sensible moral judgments held by all or nearly all philosophies and philosophers.

However, the burden seems too great, even with a division of labor, unless the panel is to be indefinitely and impractically large in terms of membership. Also, a division of labor might threaten the attainment of consensus, or even near-consensus. Furthermore, competent moral judges possessing such a

¹⁶ This scenario assumes that all sensible moral judgments of all competent moral judges are taken into account, but it does not extend to moral principles. Using moral principles to infer or derive moral judgments would violate the decision procedure. Rawls explicitly forbids the derivation of considered moral judgments from moral principles: See his second comment (1951, 180) on his criteria for competent moral judges and his seventh criterion (1951, 183) for considered moral judgments.

magnitude of knowledge in both scope and depth lose their “ordinary person” characteristics, which Rawls is eager to preserve.¹⁷

*Rawls's
dogmatism
violates his
own decision
procedure.*

To summarize: (1) The kind of dogmatism discussed here is indeed attributable to Rawls. (2) Rawls cannot truthfully deny this dogmatism. (3) Nor can he avoid the dogmatism without compromising his principles or modifying his decision procedure. (4) All this poses a systemwide threat because his own decision procedure, which does not allow dogmatism, does not justify his particular principles of justice, which can only be defended dogmatically.

Even when focusing on a single work, in this case, the first publication of Rawls, straying beyond that point is often tempting, and at times necessary, especially to avoid a possible misunderstanding at a critical juncture. This is such a juncture. Specifically, the third item in the preceding list may make one wonder whether reflective equilibrium, particularly the wide kind, might not take care of these problems.¹⁸

¹⁷ Rawls contends that competent moral judges are ordinary people in many respects. This comes out in his enumeration (1951, 178–179) of the qualifications required and in his subsequent provisions (1951, 180–181) concerning those qualifications. He maintains, for example, that “what we call ‘moral insight’ is the possession of the normally intelligent man as well as of the more brilliant” and that “a competent moral judge need not be more than normally intelligent” (1951, 178). He also defines the knowledge required of the competent moral judge in terms of what is “reasonable to expect the average intelligent man to know” (1951, 178). He claims further that he selects competent moral judges “not by means of characteristics which are the privileged possession of any race, class, or group, but which can and often do belong, at least to a certain degree, to men everywhere” (1951, 181).

¹⁸ The literature on reflective equilibrium, wide or narrow, is enormous. Perhaps true of most Rawlsian themes, it is particularly pronounced in this one, largely because other prominent philosophers have comparably prominent roles here, including Nelson Goodman (1954), who is widely credited for having inspired the concept, and Norman Daniels (1979, 1980a, 1980b, 1996), who is often said to have perfected it, not to mention many who have adopted, developed, or rejected it. While it is Rawls who coined the term “reflective equilibrium” (1971, 20–21, 48–53), and Rawls again who distinguished between “wide” and “narrow” forms of it (1974–1975, 8), the question is whether he borrowed it from Goodman’s (1954) analogous albeit unnamed approach to the problem of induction. Recently, Mäkinen and Kakkuri-Knuuttila (2013) have argued against the standard assumption of an influence on Rawls by Goodman, instead tracing the Rawlsian origin of reflective equilibrium to Rawls (1955) himself, in other words, finding room for independent development, which becomes plausible as they move the

Importing the concept of wide reflective equilibrium into the setting of the “Outline” seems like a promising response to complications arising in and about the scenario of a division of labor. These “problems” have all along been little more than playful scenarios introduced in the spirit of a devil’s advocate conducting thought experiments in a dialectical effort to sort out the justification mechanism Rawls advocates. Be that as it may, let us suppose that the wrinkles are ironed out through wide reflective equilibrium.

This would mean that we no longer need to worry about the ideological harmonization of competent moral judges, considered moral judgments, and reasonable moral principles. We would also not have to trouble ourselves with the dynamics of the reciprocal adjustments required toward equilibrium. With everything iterated into alignment, the components of the decision procedure would be in optimal operating efficiency, jointly constituting a better model of ethical reasoning than the prototype. We would, in short, be assured of a smooth operation. But the quality of the operation, so to speak, never was a threat or obstacle to begin with. The purportedly objective decision procedure is, in fact, objective. It always has been. The problem here is the dogmatism that defies the objectivity of the decision procedure already in place. Replacing that model with one boasting a more efficient justification mechanism, as promised by wide reflective equilibrium, is not going to make the dogmatism go away. It never did. That is to say, it never did in real life, as the basic design underwent change after change

Rawlsian origin from its usual spot in 1971 (*Theory*) to a new beginning in 1955 (“Two Concepts of Rules”). However, the question of outside influence remains open. Quine (1951), for example, seems just as likely a candidate as Goodman (1954). Quine’s holism, evident as early as his “Two Dogmas of Empiricism” (1951, 38), is not a far cry from reflective equilibrium. Taking this route, one may then wish to follow up on the debt of inspiration Quine acknowledges to Rudolf Carnap in the first edition of that article (1951, 38), adding Pierre Duhem in a footnote to the revised edition (1953, 41, n. 17). But why stop at Quine? A noted historian of moral and political philosophy such as Rawls would have surely felt the influence of Aristotle, who had a proclivity for reasoning in a fashion that may well count as a forerunner to the method of reflective equilibrium (*Nicomachean Ethics* 1094b13–27, 1095a31–b3, 1098a20–b8, 1098b26–29, 1109a20–b26, 1138b18–35, 1145b1–7, 1146b6–8). Indeed, Rawls himself locates the origin of the method in Aristotle, though he also professes a scholarly debt to “most classical British writers through Sidgwick” (1971, 51, n. 26).

while the moral mindset of Rawls himself did not. Rawls the moral engineer kept coming up with newer and better decision procedures for ethics, but Rawls the moral agent continued to hold the same position.

Despite numerous adjustments to his decision procedure, Rawls's liberal bias persisted.

The liberal bias has always been there whether Rawls arrived at it through a rocky process or with admirable finesse. Wide reflective equilibrium turns out to support Rawls's own moral, social, and political outlook, much in the way that his competent moral judges do, with both appeals enjoying the same convenient validation claimed by every local religion in the world as the only true religion clashing with a myriad of misguided ones. Perhaps wide reflective equilibrium promises greater objectivity than the original decision procedure. But, again, the problem has not been with the promise. The objectivity guaranteed by the original decision procedure was perfectly acceptable from the outset. It just somehow never panned out, with a succession of outcomes remaining forever skewed in favor of the moral sensibilities of Rawls: sensibilities consistent with liberalism as Rawls himself is happy to acknowledge. This, and not the operating efficiency of the decision procedure, is the real problem, at least in this portion of the present article.

The next section takes up the second problem under consideration: the possibility that the decision procedure itself may be otiose or even altogether vacuous. This problem is an extension or reflection of the first (or vice versa). The two are so closely related that the difference between them is essentially one of degree. In the first problem, the number of judges in agreement is not sufficient to justify Rawls's principles of justice. In the second, the number of judges in agreement is not sufficient to justify any moral principles at all or at least any substantive moral principles.

4. System Sterility

To recall the terminology adopted for the sake of clarity, considered moral judgments are sensible moral judgments that are acceptable to all or nearly all competent moral judges. This much is true by definition. The distinction makes it possible, without contradiction, to interpret Rawls as holding that

there is some core of sensible moral judgments (“considered moral judgments” being the wrong term here because they are already presumed stable in the manner required) transcending time, culture, and ideology, and thereby being acceptable to or accepted by any competent moral judge at any time, which is what makes those sensible moral judgments considered moral judgments.

Something similar to this appeal to a core consensus is evident in Rawls’s later works as well. Such evidence starts to show up as early as *Theory*:

Being designed to reconcile by reason, justification proceeds from what all parties to the discussion hold in common. Ideally, to justify a conception of justice to someone is to give him a proof of its principles from premises that we both accept, these principles having in turn consequences that match our considered judgments. [1971, 580–581]

The gist of this passage is that justification should proceed from premises agreed upon in advance and acceptable to all parties, hence from an objective core of moral judgments on which there is prior consensus.¹⁹ This is a plausible proposal as far as a decision procedure is concerned. But any attempt to justify Rawlsian principles of justice in this way presumes an agreement not only on basic moral intuitions or on the building blocks of ethical theory but also on substantive moral issues. Assuming a point (or region) of convergence in connection with basic moral judgments is reasonable enough, but prior consensus on answers to substantive moral questions cannot be a necessary, or even acceptable, starting point for ethical justification: It cannot be necessary, because it requires too much; it cannot be acceptable, because it begs the question of justification.

To be fair, Rawls does not take his principles of justice to

¹⁹ This is different from the later (and now more familiar) notion of “overlapping consensus,” which Rawls characterizes as a “consensus of comprehensive, or partially comprehensive, religious, philosophical, and moral doctrines” (2001, xvii) and defines as “a consensus in which the same political conception is endorsed by the opposing reasonable comprehensive doctrines that gain a significant body of adherents and endure from one generation to the next” (2001, 184). Rawls is clear about this: “The idea of an overlapping consensus was not used in *Theory*” (2001, 186). He notes further: “The term is used once, *Theory*, §59: 340, but for a different purpose than my present one” (2001, 186, n. 9).

be justified automatically with an appeal to a core of prior consensus. He finds the consensus necessary, not sufficient, for ethical justification. Still, the question is whether he forces his conception of justice into the core of prior consensus. That, indeed, would be to beg the question.

A proper answer requires a closer look at the notion of a core consensus. Questions abound: What constitutes a core of consensus? Which sensible moral judgments are included in it? Who is to choose the judgments to be included, and by what principles of selection?

The very essence of a core consensus suggests that much is left out: The core is not required to contain all the sensible moral judgments of all the competent moral judges, or even to include at least one sensible moral judgment by each competent moral judge. It is sufficient that the sensible moral judgments in the core be of a sort that would, upon consideration or reflection, be acceptable to all or nearly all competent moral judges. This is what Rawls means by "considered moral judgment," and the concept of the core brings out the distinction drawn earlier between considered moral judgments and sensible moral judgments. The core, by definition, consists of all and only considered moral judgments. Any sensible moral judgment that makes its way into the core becomes a considered moral judgment.

That said, a core of such universal appeal is unlikely to go beyond very basic judgments, and those, either severally or collectively, would have to be otiose, contributing nothing, or very little, toward the moral enlightenment of the moral agent. Sensible moral judgments in such a sterile core would be jointly insufficient to support a complete and coherent set of moral principles. In other words, the class of all considered moral judgments is either completely or very nearly devoid of useful sensible moral judgments, or it contains only unhelpful sensible moral judgments.

But what is this distinction between useful and useless (or between helpful and unhelpful) sensible moral judgments? Quite simply, the useful and helpful ones are those that make a real contribution toward the adjudication of significant moral issues, conflicts, and problems, in short, the hard cases, while the useless and unhelpful ones are those that do not fit

*Much is
necessarily left
out of a core
consensus.*

this description. Consider two specific examples: A judgment such as “it is morally wrong for parents to have their children executed by hired assassins to get them out of the house” is probably one that would be acceptable to all or nearly all competent moral judges, and it represents the kind of judgment that could be found in an objective core, but it is useless and unhelpful in the way just described. On the other hand, a judgment such as “it is morally wrong for a woman to have the fetus in her womb killed and extracted by medical personnel to terminate the pregnancy” is probably not one that would be acceptable to all or nearly all competent moral judges, and it is not one that could make its way into the objective core, yet it is the kind of sensible moral judgment that captures the issues that actually matter to us, whether as moral philosophers or as ordinary persons.

In the same vein, useful and helpful judgments are substantive, while useless and unhelpful ones tend to be trivial. It may be objected, on behalf of Rawls, that requiring substantive moral judgments to be included in the core to make it useful, while at the same time prohibiting them from being included so as not to beg the question, is to create an artificial dilemma for Rawls, one for which he is not accountable. It may thus be argued that where the question-begging really takes place is in opposing Rawls in this way, that is, in reducing considered moral judgments to trivial moral judgments. It may be objected, moreover, that the distinction between what may and may not be admitted into the objective core is biased against Rawls, making his decision procedure look circular when it really is not, and that the proper distinction is not between “trivial” and “substantive” moral judgments but between “basic” and “derivative” moral judgments.

Such persistent protest calls for consideration. The two examples just given, so the objection goes, are both irrelevant because they are both derivative. Neither one can be included in the core, because the core contains only basic judgments. Both the judgment that “it is morally wrong for parents to have their children executed by hired assassins to get them out of the house” and the judgment that “it is morally wrong for a woman to have the fetus in her womb killed and extracted by medical personnel to terminate the pregnancy” are derived

from a basic moral judgment such as “murder is wrong” (or to add specificity, a judgment such as “it is wrong to kill people unjustly”). This basic judgment can be included in the core, but the two derivative ones cannot. This is because, even where everyone agrees on the basic judgment, it does not automatically entail the two derivative judgments. The possibilities are endless in manipulating a basic judgment to derive others, even ones that contradict one another.

The more one enriches the set of moral judgments on which prior consensus is presumed, the less likely actual agreement becomes.

What follows from this objection? Suppose that the distinction between trivial and substantive moral judgments is, in fact, irrelevant, and that the appropriate distinction is between basic and derivative moral judgments. If only basic sensible moral judgments can be admitted into the core, then all considered moral judgments must be basic ones. Then so much the worse for Rawls. Basic moral judgments can no more justify a moral principle than can trivial moral judgments. Calling it something else does not change the nature of the judgment. One might imagine unanimous or widespread agreement on a great many trivial or basic moral judgments, and one might imagine unanimous or widespread agreement on very few substantive or derivative moral judgments, but neither kind of agreement would yield a core rich enough to provide anything like a complete set of moral principles. On the other hand, one could also imagine agreement on a plethora of substantive or derivative sensible moral judgments, but then the agreement would not be unanimous or even widespread. The more one enriches the set of moral judgments on which prior consensus is presumed, the less likely actual agreement becomes. And it does not matter whether the enrichment is in number, significance (trivial vs. substantive), or level (basic vs. derivative).

The question is whether Rawls requires only a basic kind of agreement necessary to any moral theory or assumes too much to begin with. When the starting point takes many if not most disagreements in attitude to be settled before theory construction begins, the resulting theory will be circular, justifying what people already believe to be right or wrong. The issue turns on whether a core general enough to reflect unanimity, or a very high level of agreement, can at the same time be rich enough for the set of sensible moral judgments it contains to play the crucial role Rawls envisions it to play in ethical

justification. If the core is general enough to secure a respectable consensus, it just may be too general to facilitate theory construction. If it is genuinely helpful in theory construction, it may not be general enough to count as a consensus.

To recapitulate, pausing for a provisional conclusion, the answer to whether Rawls's decision procedure is vacuous or otiose is that it is not vacuous, since it can, in any event, justify some moral principle or other, but that it is dangerously close to being otiose, because it cannot reasonably be held to justify very many substantive moral principles.

A final objection on behalf of Rawls may be that the critique confuses two different things, a prior consensus and a final consensus: The passage cited above from *Theory* (1971, 580–581) requires a core of *prior consensus* as a starting point in ethical justification, whereas the agreement of competent moral judges on considered moral judgments, as depicted in the "Outline," is not a prior consensus but a *final consensus*, which follows from, and is built upon, deliberations and modifications on a prior consensus, much like the one described in *Theory*. Thus, the objection runs, while it is true that the core of prior consensus, from which theory justification begins, can include only basic moral judgments, it is permissible to include derivative moral judgments (and even substantive moral judgments) in the core of final consensus.

This is true but irrelevant. The objection brings up a distinction which, in its proper sense, has in fact been observed by the present critique.²⁰ The relevant consensus, whether prior or final, is about moral judgments, whereas the justification promised by the decision procedure is about moral principles. The only essential feature of a "prior" consensus (always on moral judgments) is that it precedes the justification process (always of moral principles). That is the sense of priority intended by the term (the level of complexity being an accidental feature). A "final" consensus, in contrast, is the agreement prevailing

²⁰ This is still not about the question of an "overlapping consensus," which is a more specific concern than the generic distinction between a prior and final consensus. See n. 19 above. Any talk of consensus, however, particularly with rigorous distinctions, is a natural reminder of Rawls's "overlapping consensus." If any association were to be attempted at this point, it would be safer to say that the overlapping consensus is closer to what is depicted here as a final consensus.

after justification. There is no room for confusing the two. The point of the critique here (in this section) is that the decision procedure Rawls recommends may possibly be operationally sterile in the sense of persistently lacking the level of consensus required for the kind of justification proposed. And the practical counterpart of this systemic problem, as discussed in the previous section, is that the justification Rawls needs for the principles he advocates requires too much prior agreement on moral judgments.²¹

How, then, does Rawls arrive at his conception of justice as fairness? Is it the natural result of his decision procedure for ethics? Or does he manipulate the system, if ever so slightly, starting with a core of prior consensus that has a sufficiently liberal touch for the principles of justice to come out just right? The evidence is not conclusive either way, but the second possibility presents a nagging uncertainty. To take any moral principles as rich as those that represent Rawls's conception of justice and to push them (or whatever input yields that output) into a core of consensus is to beg the question. Strictly speaking, it is meaningless to speak of pushing any moral principles at all, rich or otherwise, into a core consensus, which is defined as containing only moral judgments and not moral principles. But what is meant, at any rate, is rigging the system in some way, whether this be done with principles or with judgments.

²¹ A more meaningful objection, though still not a conclusive one, would be that the prior consensus can be strong and rich enough, without including derivative or substantive moral judgments, to help adjudicate the tough cases and to formulate complex moral principles. The consensus envisaged here would presumably proceed from a minimal base of agreement sufficient for arriving at a greater base of agreement, a process gradually unfolding toward an eventual meeting of the minds on important matters. But this, again, seems to be wishful thinking. Assumptions for the sake of argument are perfectly acceptable in thought experiments, but once the arguments are laid out, the conclusions are no stronger than the assumptions. There is no such consensus. We do disagree on abortion, on gun control, on genetic engineering, and so on. Evidently, whatever moral convergence we draw on to decide such matters is not a strong enough base of agreement to bring us together on such matters.

5. Broader Context

Are the foregoing problems restricted to the “Outline” of 1951? Despite a primary focus on the “Outline,” largely for the sake of specificity, I have been trying, where possible and relevant, to bring out connections with Rawls’s later works. A limited effort of that sort is no substitute for a comprehensive survey. While there may (or may not) be room to object that Rawls’s later thought is free of the problems discussed here, such an objection cannot fairly be pursued as a speculative point of protest without concrete evidence. That effort would also have to be specific, not to mention comprehensive, thus preferably drawing on a single theme ranging over multiple works. A good example—both specific and comprehensive—is that of Rawls’s pluralism.

The objection in that context would be that saddling Rawls with dogmatism is wrong in view of his later move to embrace pluralism (e.g., 1987, 18–22; 1993, 144–168; 2001, 3–5). But this pluralistic turn is not necessarily a turn away from dogmatism. Even as he embraces pluralism, for example, in a later paper on “overlapping consensus” (1987, 1–25)—before the second edition of *Theory* (1999) and before either edition of *Political Liberalism* (1993/2005)—his thoughts turn to a liberal conception of justice: “As I have said, the most reasonable political conception of justice for a democratic regime will be, broadly speaking, liberal” (1987, 17).²²

*Pluralistic
turn not
necessarily
a turn away
from
dogmatism.*

Pluralism is indeed a common thread running through much of Rawls’s mature output. The theme of pluralism, however, goes hand-in-hand with that of an “overlapping consensus” (2001, 32–38, 192–195), which, in turn, works in favor of political liberalism. Diluting pluralism through an appeal to an overlapping consensus (which conveniently happens to favor a certain type of outcome) is no better than ignoring or rejecting it through an appeal to competent moral judges (who consistently happen to favor a certain type of outcome). The competent moral judges are replaced by “reasonable persons” who are supposed to see in time, and typically over generations, that a conception of “justice as fairness” (which Rawls

²² This liberal conception of justice is explored throughout the work in question (1987, 1–25) and articulated specifically on p. 18, including especially n. 27. See 1993, 3–46, for a fuller exposition.

identifies in his later works as a “liberal conception of justice”) is an ideal they all hold in common, despite any ideological differences they may otherwise continue to have. This is the overlapping consensus invoked in justification.²³

This gradual realization is a manifestation of the “reasonable moral psychology” (e.g., 1987, 22, 23) expected of everyone (specifically, of every citizen in a constitutional democracy). Rawls often appeals to moral psychology, whether or not he uses that exact term:

[1] Note also that the success of liberal institutions may come as a discovery of a new social possibility: the possibility of a reasonably harmonious and stable pluralist society. [1987, 23]

[2] The kind of stability required of justice as fairness is based, then, on its being a liberal political view, one that aims to be acceptable to citizens as reasonable and rational, as well as free and equal, and so as addressed to their public reason. [2001, 185]

And he uses that exact term in several places, including, for example, *Theory* (1971, 490–496 [=1999, 429–434]), the “Reply to Habermas” (1988, 270), and *Political Liberalism* (1993, 86–88, cf. 158–168). Moreover, he qualifies the psychology in question as a “reasonable” one, thus referring explicitly to a “reasonable moral psychology,” among other places, in his treatment of “overlapping consensus” (1987, 22, 23) and in his explication of “justice as fairness” in the *Restatement* (2001, 195–198).²⁴

This is a reflection of Rawls’s general tendency to claim “reasonableness” for the key ideas, premises, and assumptions in his moral, social, and political outlook.²⁵ While we all want our ideas, arguments, and positions to be reasonable, calling

Rawls tends to claim “reasonableness” for his key ideas and assumptions.

²³ Rawls provides succinct definitions (2001, xvii, 184) of “overlapping consensus” as well as detailed discussions of it (1985, 225–226; 1987, 1–25; 1993, 133–172; 2001, 32–38, 192–195). See n. 19 above for one such definition.

²⁴ In *Theory* (1971), his appeals are variously to the “facts” (pp. 143, 176), “laws” (p. 177), and “principles” (pp. 490–496) of moral psychology, and ultimately to a “general knowledge” (p. 176) of it.

²⁵ Even a cursory survey of a single work (2001) shows a trend: “reasonable comprehensive doctrine” (pp. 34, 183, 184); “reasonable conception” (p. 37); “reasonable disagreement” (p. 35); “reasonable doctrine” (p. 37); “reasonable human psychology” (p. 184); “reasonable moral psychology” (p. 181); “reasonable overlapping consensus” (pp. 32–38); “reasonable person” (pp. 32–38); “reasonable pluralism” (pp. 32–38).

something “reasonable” does not make it so. No doubt, this is not what Rawls thinks either, but he does introduce reasonableness, in each case, as an assumption, from which, together with other premises, he draws certain conclusions. And the conclusions are invariably in favor of liberalism. Given that arguments are meant to support convictions, all the arguments adduced for a certain position will naturally be in favor of that position, but the support becomes questionable when it rests on postulated reasonableness. Rawls works with a significant correlation (far greater than would be expected of random outcomes with independent variables) between the reasonable and the liberal, whereby specifying reasonableness as a formal criterion somehow always points to a liberal outcome.

Calling something “reasonable” does not make it so.

It may be objected that Rawls is using “reasonable” in a special sense, anchored to his constructivism, such that “reasonable” is meant not so much to convey reasonableness in the ordinary sense as it is to eschew “truth” in the intuitionist sense (1980, 554). He would thus be contrasting “reasonable” with “true” rather than with “unreasonable.” That would be par for the course in a metatheoretical position, but there is neither a need nor any room in this context for such relentless pursuit of that distinction. Rawls invokes the qualification “reasonable” at every step of theory construction and theory justification, as well as in discussing corollaries, implications, and applications. Outside metatheoretical discussions, he should have no greater recourse to the reasonable than the intuitionist has to the true. I have yet to come across an example of an intuitionist working out descriptive or normative theories, or any applications thereof, by appealing to so many variables that are simply true. But I do see Rawls doing this with variables that are simply reasonable.

The emphasis on the reasonable may alternatively, or in addition, be read as a reflection of Rawls’s intellectual modesty. After all, Rawls never claims that the considered moral judgments of competent moral judges are explicated only by the seven moral principles he lays out in his “Outline” (1951, 192–193). He proposes only that such judgments are, in fact, explicated by those principles, leaving open the question whether they are or can be explicated by others as well. Actually, the question is not as open as it may seem, considering that suc-

cessful consultation with the judges requires unanimous (or near-unanimous) agreement, which, in turn, leaves little to no room for multiple outcomes. Nevertheless, it must be granted that Rawls is emphatically tentative in his presentation of the original principles. He characterizes the set of principles as “a statement of what are hoped to be satisfactory principles of justice” and goes on to qualify that statement as “provisionary” (1951, 191).

This might have been a compelling objection had such intellectual modesty come with room for alternatives. It does not. Rawls leaves no such room in the “Outline,” where he works with the assumption “that a satisfactory and comprehensive explication of the considered judgments of competent judges is already known” (1951, 187). What is relevant here is not that the said judgments are already known, which I grant as an assumption, but that the explication is both satisfactory and comprehensive with respect to the considered moral judgments of competent moral judges. This suggests that Rawls, despite his modesty, does not mean to leave open the question regarding the tenability of competing moral platforms. If his seven moral principles (or any later reformulations in any combination) are satisfactory and comprehensive with respect to the considered moral judgments of competent moral judges, this leaves no room for a competing moral framework. A satisfactory and comprehensive explication of the very same data set (the considered moral judgments of competent moral judges) cannot be had through competing (mutually inconsistent) perspectives.²⁶ Any plurality would have to be limited to redundant systems (or redundant “reasonable comprehensive doctrines”). And this means that Rawls places far greater confidence in his own principles (and far less in others) than his

²⁶ The emphasis on “comprehensive” here is not yet the technical sense established later (1985) to distinguish between a “political” conception and a conception derived from a “comprehensive” doctrine—a general distinction also instantiated as a specific contrast between “political” and “metaphysical” conceptions of justice (e.g., 1985, 1987, 1988). See, for example, the *Restatement* (2001) of justice as fairness, where Rawls acknowledges that the relevant distinction had not yet been brought to bear in *Theory* (nor therefore prior to it): “That work [*Theory*] never discusses whether justice as fairness is meant as a comprehensive moral doctrine or as a political conception of justice” (2001, 186).

identification of them as provisional would seem to suggest.

Truth be told, I am comfortable with Rawls's conception of justice, including all the principles explicating it. But this is a confession, not an argument. His principles of justice, in any formulation, have no inherent advantage over the principles of other "reasonable" systems, say, utilitarian or deontological ones, in terms of justificatory support from his own decision procedure for ethics. Yet his principles are so basic that they might even be their own justification. And if they are not themselves foundational, they might still pass for a derivative of the categorical imperative, perhaps even for a direct albeit complex restatement of it. In either case, no justification would seem to be required if by that is meant proof in the sense famously rejected by John Stuart Mill.²⁷ And perhaps none is possible (as Mill insists).

That is why, when all is said and done, the ultimate justification Rawls leaves us with is that a liberal conception of justice is something of an acquired taste, or as he puts it, an "acquired allegiance" (e.g., 1987, 21–22), a habituation of sorts that gradually yet persistently makes its way into the sociocultural fabric of political existence (1993, 158–168; 2001, 192–195). His account of the possibility of overlapping consensus as the foundation of political liberalism rests, in the end, on this appeal to collective habituation—a part of the "reasonable moral psychology" (e.g., 1987, 22, 23) expected of everyone (i.e., of every citizen).

Rawls's ultimate justification: that a liberal conception of justice is an "acquired allegiance."

This is why I am satisfied with his principles but not with his justification of them, that is, neither with his justification in the "Outline" (which is what is at the forefront here) nor with his justification elsewhere (which is just as relevant). The reason is not that I have been able to develop a better justification model myself or that I have adopted one already developed by someone else. I should not have to do either as a condition for rejecting what does not work. My failure is not Rawls's success.

There is no compelling indication that Rawls's principles of justice are justified by his decision procedure in any sense or to

²⁷ This is Mill's position on the nature of proof in questions of ultimate ends. He discusses it in *Utilitarianism* (1969, 207–208, 234): chapter one, paragraph five; chapter four, paragraph one.

any degree that sets them apart from other actual or conceivable positions on justice whose principles may perform just as well in the same decision procedure. As for the later appeal to what will naturally emerge through many generations as the most palatable conception of justice, this is entirely possible. But even there we do not have much more than Rawls's considered opinion that the overlapping consensus will favor a liberal conception of justice, specifically the one formulated by Rawls, as opposed to a conservative alternative, or really any alternative whatsoever.

On the whole, Rawls is as open to alternatives as is humanly possible, which does not rule out a certain degree of rigidity in his political views. As he later acknowledges openly (*The Law of Peoples*), he is neither enthusiastic nor optimistic about dialogue with people who are not on board with political liberalism:

Those who reject constitutional democracy with its criterion of reciprocity will of course reject the very idea of public reason. For them the political relation may be that of friend or foe, to those of a particular religious or secular community or those who are not; or it may be a relentless struggle to win the world for the whole truth. Political liberalism does not engage those who think this way. [1999, 132]

This seems to leave plenty of room for dogmatism, and not as much as promised for pluralism. An overlapping consensus will, of course, be liberal if alternatives are excluded in the first place. It is difficult to see how Rawls's later works might readily absolve him of the application bias identified in this article.

Works Cited

- Aristotle. 1984. *Nicomachean Ethics*. Translated into English by William David Ross, revised by James Opie Urmson, and published as part of *The Complete Works of Aristotle: The Revised Oxford Translation* (pp. 1729–1867). Second of two volumes. Edited by Jonathan Barnes. Princeton: Princeton University Press.
- Daniels, Norman. 1979. "Wide Reflective Equilibrium and Theory Acceptance in Ethics." *The Journal of Philosophy* 76:5 (May): 256–282.

-
-
- Daniels, Norman. 1980a. "On Some Methods of Ethics and Linguistics." *Philosophical Studies* 37:1 (January): 21–36.
- Daniels, Norman. 1980b. "Reflective Equilibrium and Archimedean Points." *Canadian Journal of Philosophy* 10:1 (March): 83–103.
- Daniels, Norman. 1996. *Justice and Justification: Reflective Equilibrium in Theory and Practice*. Cambridge: Cambridge University Press.
- Goodman, Nelson. 1954. *Fact, Fiction, and Forecast*. London: University of London, The Athlone Press. Numerous editions and reprints, the last in the author's lifetime being the fourth edition (Cambridge: Harvard University Press, 1983).
- Hart, Herbert Lionel Adolphus. 1973. "Rawls on Liberty and Its Priority." *The University of Chicago Law Review* 40:3 (Spring): 534–555.
- Mäkinen, Jukka, and Marja-Liisa Kakkuri-Knuuttila. 2013. "The Defence of Utilitarianism in Early Rawls: A Study of Methodological Development." *Utilitas* 25:1 (March): 1–31.
- Mill, John Stuart. 1969. *Utilitarianism*. Referenced in this article as it appears in *Essays on Ethics, Religion and Society* (pp. 203–259), vol. 10 of the *Collected Works of John Stuart Mill*. Edited by John Mercel Robson. Series editor, F. E. L. Priestly. Toronto: University of Toronto Press. Originally published as an essay in three parts in *Fraser's Magazine for Town and Country*: chs. 1–2 in *Fraser's Magazine* 64:382 (October 1861): 391–406; chs. 3–4 in *Fraser's Magazine* 64:383 (November 1861): 525–534; ch. 5 in *Fraser's Magazine* 64:384 (December 1861): 659–673.
- Quine, Willard Van Orman. 1951. "Two Dogmas of Empiricism." *The Philosophical Review* 60:1 (January): 20–43. Revised and reprinted as chapter 2 of Quine's *From a Logical Point of View: Nine Logico-Philosophical Essays* (pp. 20–46). Cambridge: Harvard University Press, 1953. The book was revised and released in a second edition in 1961, and reprinted with a new foreword in 1980, but the emendations and revisions noted in the front matter in either case do not include the essay itself.
- Rawls, John. 1942. *A Brief Inquiry into the Meaning of Sin and Faith: An Interpretation Based on the Concept of Community*.

-
-
- Senior Thesis (Princeton University). Bundled together with a short manuscript from Rawls's maturity ("On My Religion") and published posthumously as a book (Cambridge: Harvard University Press, 2009).
- Rawls, John. 1950. *A Study in the Grounds of Ethical Knowledge: Considered with Reference to Judgments on the Moral Worth of Character*. Ph.D. Thesis (Princeton University). See *Dissertation Abstracts* 15 (1955): 608–609.
- Rawls, John. 1951. "Outline of a Decision Procedure for Ethics." *The Philosophical Review* 60:2 (April): 177–197.
- Rawls, John. 1955. "Two Concepts of Rules." *The Philosophical Review* 64:1 (January): 3–32.
- Rawls, John. 1958. "Justice as Fairness." *The Philosophical Review* 67:2 (April): 164–194. Revised and expanded edition of a paper delivered at the Eastern Division meeting of the American Philosophical Association (Harvard University, 27–29 December 1957) shortly after publication in *The Journal of Philosophy* 54:22 (24 October 1957): 653–662.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge: Harvard University Press. Second ("revised") edition, 1999. References are primarily to the original, and sometimes to both, but never to the second alone. A date is supplied in either case to avoid confusion.
- Rawls, John. 1974–1975. "The Independence of Moral Theory." *Proceedings and Addresses of the American Philosophical Association* 48: 5–22. Presidential Address delivered before the seventy-first annual meeting of the American Philosophical Association, Eastern Division, Washington, 28 December 1974.
- Rawls, John. 1980. "Kantian Constructivism in Moral Theory." *Journal of Philosophy* 77:9 (September): 515–572.
- Rawls, John. 1982. "The Basic Liberties and Their Priority." Tanner Lectures on Human Values 3. Salt Lake City: University of Utah Press. Revised and expanded version of lecture delivered by Rawls on 10 April 1981 at the University of Michigan. Reprinted as lecture 8 in *Political Liberalism* (1993, 289–371).
- Rawls, John. 1985. "Justice as Fairness: Political Not Metaphysical." *Philosophy & Public Affairs* 14:3 (Summer): 223–251.

-
-
- Rawls, John. 1987. "The Idea of an Overlapping Consensus." *Oxford Journal of Legal Studies* 7:1 (Spring): 1–25.
- Rawls, John. 1988. "The Priority of Right and Ideas of the Good." *Philosophy & Public Affairs* 17:4 (Autumn): 251–276.
- Rawls, John. 1993. *Political Liberalism*. New York: Columbia University Press. Second ("expanded") edition, 2005. References here are to the original.
- Rawls, John. 1999. *The Law of Peoples: With "The Idea of Public Reason Revisited."* Cambridge: Harvard University Press.
- Rawls, John. 2001. *Justice as Fairness: A Restatement*. Edited by Erin Kelly. Cambridge: Harvard University Press.