



John Rawls and the Search for Stability

A Theory of Justice by John Rawls; Political Liberalism by John Rawls

Review by: Brian Barry

Ethics, Vol. 105, No. 4 (Jul., 1995), pp. 874-915

Published by: [The University of Chicago Press](#)

Stable URL: <http://www.jstor.org/stable/2382115>

Accessed: 27/12/2014 10:44

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *Ethics*.

<http://www.jstor.org>

John Rawls and the Search for Stability

*Brian Barry**

I. INTRODUCTION

The original impetus for this essay was provided by reviewers of my book *Theories of Justice*.¹ There I had at a number of points discussed the ideas of John Rawls, but had confined myself to *A Theory of Justice*.² Almost without exception, I was taken to task for not having discussed Rawls's later views. In some cases, it seemed to be implied that Rawls's change of mind somehow invalidated his earlier ideas. The subsequent reaction to *Political Liberalism* has confirmed my suspicion that Rawls is widely assumed to have got one thing right at least: that there was something wrong with *A Theory of Justice* that needed fixing.³ However unconvinced they may be by the positive doctrine, Rawls's commentators have tended to accept uncritically his own disparagement of *A Theory of Justice*.

It is my contention that Rawls's account of what was wrong with *A Theory of Justice* will not withstand scrutiny. If I am right about this (and I argue the case in the next section), an intriguing question opens up. We have to ask, What does account for Rawls's shift to the

* The main ideas in this article were already worked out several years before the publication of Rawls's recent book and the commentaries on it discussed here. I have been sharing my perplexities from the beginning with Matt Matravers. I should like to acknowledge his help in this regard and also in the preparation of the final draft. I am also very grateful to Percy Lehning for his generosity in giving me detailed written comments on earlier drafts, drawing on his extraordinary knowledge of the Rawls canon and the secondary literature. Some of the ideas were tried out at University College, Dublin; Trinity College, Dublin; and Cambridge University. Discussion on each of these occasions has been valuable in suggesting where the argument needed more work.

1. Brian Barry, *Theories of Justice* (Berkeley: University of California Press; Hemel Hempstead: Harvester-Wheatsheaf, 1989).

2. John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), hereafter cited in the text as *TJ*.

3. John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), hereafter, cited in text as *PL*.

Ethics 105 (July 1995): 874–915

© 1995 by The University of Chicago. All rights reserved. 0014-1704/95/0504-4009\$01.00

positions developed in the last decade and now brought together in *Political Liberalism*? The only doubts that have been expressed about Rawls's own account of his intellectual odyssey concern his denial (*PL*, p. xvii, n. 6) that he has changed his views in response to "communitarian" criticisms. I accept this denial, however, and I accept his claim that everything distinctive about *Political Liberalism* stems from a concern with the stability of justice (*PL*, pp. xv–xvi). This suggests that we need to get clear about what Rawls takes the problem of stability to be. Section III is therefore addressed to the problem of stability, which is understood in much the same way in both *Political Liberalism* and *A Theory of Justice* as turning on the existence of appropriate motivation for doing what justice requires.

If *Political Liberalism* is occasioned by unresolved problems about stability left over from *A Theory of Justice*, the obvious next step is to see what Rawls has to say about it there. Sections IV–VI are designed to do that, focusing on the last chapter, which makes stability turn on "the good of justice." I argue that, when we see what is needed for stability and what stability is needed for, we can understand why Rawls became dissatisfied with his account of the conditions of stability in *A Theory of Justice*.

With Section VII, we make the transition to *Political Liberalism*. This section is concerned with the argument that Rawls suggests can be made for what he calls a "political conception" of justice. This argument is intended to be "free-standing," independent of any "comprehensive view." I put the case for thinking that, if this argument is valid, it is sufficient by itself to ensure the kind of stability Rawls is interested in.

Rawls, however, maintains that the stability of justice can be assured only if the principles of justice derived within the "political conception" can be shown to be an element in an "overlapping consensus" of "reasonable comprehensive views." Having determined (in Sec. VIII) what makes a "comprehensive view" reasonable, I argue that Rawls runs together two different stories about the relation between reasonableness and justice, one about reasonable people and the other about reasonable views. This contention is developed in Section IX, which takes up one issue discussed by Rawls. This is: What can be said to someone who maintains that "outside the church is no salvation"? I show that Rawls gives two contradictory answers. One says that this claim has to be rejected because it is incompatible with "justice as fairness." The other says that it is not necessary to reject it: all we have to do is deny that it is reasonable to seek to base a political order on it. This is in effect an affirmation of the self-sufficiency of the "political conception." If it is a valid move, it obviates the need for an "overlapping consensus."

Despite this conclusion, I ask in the next two sections if there is more to be said. Can the necessity of "overlapping consensus" for

stability be reinstated in some form? In Section X, I consider the possibility that the weak point in the “political conception” might be thought to be its reliance on the “fact of pluralism,” understood (as it has to be to make the argument go) not as a brute fact but as a claim about the limits of reason. If this is too “metaphysical” (and therefore too controversial) to play a role in an argument addressed to all reasonable people, regardless of their “comprehensive views,” we are back to “overlapping consensus” as a condition of stability. Section XI identifies a form of stability—“liberal democratic stability”—which is incompatible with widespread rejection of the moral basis of the regime, whether this rejection is reasonable or not. Concern for this kind of stability would lead us to an interest in the prospects for “overlapping consensus,” in a way that I explain.

These two sections justify some attention to Rawls’s candidates for inclusion in the “overlapping consensus.” Five potential members are identified and examined in Section XII. The following section is devoted to an assessment of Rawls’s success in seeking to show that the “political conception” is compatible with each of the candidates. Section XIV draws the article to a close with a discussion of the place of *Political Liberalism* in Rawls’s work and an estimate of Rawls’s long-run position as a political philosopher.

II. RAWLS ON RAWLS

Writing in this journal, Marcus Singer told a story about a paper he had written in which he had repudiated a view with which he had formerly been associated. Alan Donagan, whose task it was to reply to this paper, prefaced his response with the remark that “a philosopher is seldom more interesting than when he criticizes his own former views; and a critic is seldom in greater peril than when he proposes to succour a colleague’s forsaken brain-children.”⁴ My own enterprise exhibits even greater temerity. For, not content with defending Rawls’s earlier self against his later self, I wish to maintain that the later Rawls attributes to the Rawls of *A Theory of Justice* errors that he did not in fact commit.

What exactly was wrong with *A Theory of Justice*, according to Rawls? In a nutshell, his answer is that the theory contained in the earlier book constituted a “comprehensive philosophical doctrine.” And this has the implication that it cannot form the charter—the basis of agreement on fundamentals—of a modern society. For any such society “is characterized . . . by a pluralism of incompatible yet reasonable comprehensive doctrines. No one of these doctrines is affirmed

4. Marcus G. Singer, “Alan Donagan: Some Reminiscences,” *Ethics* 104 (1993): 135–42, p. 137.

by citizens generally. Nor should one expect that in the foreseeable future one of them, or some other reasonable doctrine, will ever be affirmed by all, or nearly all, citizens" (*PL*, p. xvi).

Why should we believe that *A Theory of Justice* contained a "comprehensive philosophical doctrine"? Rawls concedes that "the distinction between a political conception of justice and a comprehensive philosophical doctrine is not discussed" in *A Theory of Justice*. He says nevertheless that "once the question is raised it is clear . . . that the text regards justice as fairness [i.e., Rawls's own theory of justice] and utilitarianism [conceived as the main rival to it] as comprehensive, or partially comprehensive, doctrines" (*PL*, p. xvi). I have to say that, as far as his own theory is concerned, it is far from clear to me.

A doctrine is comprehensive, Rawls tells us, "when it includes conceptions of what is of value in human life, as well as ideals of personal virtue and character, that are to inform much of our non-political conduct (in the limit our life as a whole)" (*PL*, p. 175). It may be recalled that Rawls has hedged his bets by saying that his own theory and utilitarianism were "comprehensive or partially comprehensive doctrines." We therefore need also to know what characterizes a "partially comprehensive doctrine." According to Rawls, then, the distinction is as follows. "A doctrine is fully comprehensive when it covers all recognized values and virtues within one rather precisely articulated scheme of thought; whereas a doctrine is only partially comprehensive when it comprises certain (but not all) nonpolitical values and virtues and is rather loosely articulated" (*PL*, p. 175). Thus, a partially comprehensive doctrine is the same kind of thing as a fully comprehensive doctrine—a point that Rawls himself emphasizes by adding that "by definition, for a conception to be even partially comprehensive, it must extend beyond the political and include nonpolitical values and virtues" (*PL*, p. 175).

If this is what a (fully or partially) comprehensive doctrine is, why does Rawls believe that *A Theory of Justice* contains one? We are not offered a lot in the way of elucidation. The essential point seems to be that "no distinction is drawn between moral and political philosophy," so that "a moral doctrine of justice general in scope is not distinguished from a strictly political conception of justice" (*PL*, p. xv). However, Rawls himself insists that, while justice as fairness is a political conception, it is also a moral conception. The political is thus not to be contrasted with the moral. Rather, the political is a subset of the moral, defined by its limited subject matter. Thus, Rawls says that "while [a political] conception [of justice] is, of course, a moral conception, it is a moral conception worked out for a specific kind of subject, namely for political, social, and economic institutions" (*PL*, p. 11).

On this definition of the political, it is hard to see why *A Theory of Justice* should not be said to contain a political conception of justice. *A Theory of Justice* is a moral theory, inasmuch as justice is an aspect

of morality; but it does not include “conceptions of what is of value in human life, as well as ideals of personal virtue and character” (*PL*, p. 175). The whole point of *A Theory of Justice* was that it left people to form, revise, and pursue their own conceptions of the good. (That is why the agents in the original position did not know their conception of the good: they would then have no incentive for building into the principles of justice a preference for any particular conception of the good.) Rawls himself put the crucial point as follows:

In a well-ordered society citizens hold the same principles of right and they try to reach the same judgment in particular cases. . . . On the other hand, individuals find their good in different ways, and many things may be good for one person that would not be good for another. Moreover, there is no urgency to reach a publicly accepted judgment as to what is the good of particular individuals. . . . In a well-ordered society, then, the plans of life of individuals are different in the sense that these plans give prominence to different aims, and persons are left free to determine their good. [*TJ*, pp. 447–48]

This surely makes it clear that the theory of justice is compatible with a variety of “comprehensive views”—if we take each conception of the good to give rise to at least one “comprehensive view.”

Rawls set out his grounds for dissatisfaction with *A Theory of Justice* more fully in an article published a few years before *Political Liberalism*.⁵ Here he observes that he said in *A Theory of Justice* “that if justice as fairness succeeds reasonably well, a next step would be to study the more general view suggested by the name ‘rightness as fairness.’”⁶ He maintains on the basis of this that “the reader might reasonably conclude . . . that justice as fairness is set out as part of a comprehensive view that may be developed later were success to invite.”⁷ But it seems to me that only a very obtuse reader could conclude any such thing. Surely a Rawlsian theory of “rightness as fairness” would be of the same nature as the theory of “justice as fairness.” That is to say, it would likewise set out ground rules designed to enable people with different conceptions of the good to live together without conflict in a society. The only difference would lie in its subject matter, which would consist of moral rules rather than the institutions making up the “basic structure of society.” But it would similarly prescribe limits to permissible action rather than setting out substantive ends.

5. John Rawls, “The Domain of the Political and Overlapping Consensus,” *New York Law Review* 64 (1989): 233–55, see the sec. “Comparison with *A Theory of Justice*,” pp. 248–49. Hereafter, cited as “The Domain of the Political.”

6. *Ibid.*, p. 248, referring to *A Theory of Justice*, p. 17.

7. *Ibid.*

Rawls says himself in the passage from *A Theory of Justice* already quoted that “in a well-ordered society citizens hold the same principles of right”—and he affirms that this is consistent with their being “left free to determine their good” (*TJ*, pp. 447–48; my emphasis). A Rawlsian theory of “rightness as fairness” developed as an extension of “justice as fairness” would thus be concerned with the right rather than the good. In Rawls’s terms, it would be a political conception. It would also be a moral conception—but only in the same sense as that in which Rawls’s political conception of justice is a moral conception.

We have seen that in *A Theory of Justice* Rawls makes use of the idea of a “well-ordered society” to make his contrast between the right (on which agreement is required) and the good (on which it is not). Curiously, however, he invokes precisely this concept in *Political Liberalism* in the course of explaining what he now thinks was wrong with *A Theory of Justice*. Thus, he says that the “fact of a plurality of reasonable but incompatible comprehensive doctrines—the fact of reasonable pluralism—shows that, as used in *Theory*, the idea of a well-ordered society of justice as fairness is unrealistic” (*PL*, p. xvii). This defect in *A Theory of Justice* arises, according to Rawls, because in that book “an essential feature of a well-ordered society associated with justice as fairness is that all its citizens endorse this conception on the basis of what I now call a comprehensive philosophical doctrine” (*PL*, p. xvi). However, if we compare what Rawls says about the specification of a well-ordered society in *Political Liberalism* with what he said in *A Theory of Justice*, we find that they are identical.

The section of *Political Liberalism* devoted to “The Idea of a Well-Ordered Society” (*PL*, pp. 35–40) gives three criteria. First, it is “a society in which everyone accepts, and knows that everyone else accepts, the very same principles of justice.” Second, “its basic structure . . . is publicly known, or with good reason believed, to satisfy these principles. And, third, its citizens have a normally effective sense of justice and so they generally comply with society’s basic institutions, which they regard as just” (*PL*, p. 35). In *A Theory of Justice*, a well-ordered society is defined as one “in which everyone accepts and knows that the others accept the same principles of justice, and the basic social institutions satisfy and are known to satisfy these principles.” It is “also regulated by its public conception of justice [which] implies that its members have a strong and normally effective desire to act as the principles of justice require” (*TJ*, pp. 453–54). Verbal details aside, these are manifestly the same three criteria. There is no support for Rawls’s current claim that a well-ordered society as conceived in *A Theory of Justice* required the citizens to endorse the principles of justice on the basis of a comprehensive philosophical doctrine.

Rawls himself has offered only one reason for thinking that *A Theory of Justice* was not intended to be a “comprehensive conception.”

This is that in *A Theory of Justice* “our relations to other living things and to the natural order” were explicitly excluded.⁸ This seems to me to be a minor consideration. The crucial reason for saying that *A Theory of Justice* was not “comprehensive” is that it left it open to people to determine and pursue their own conceptions of the good. It offered fair terms on which people with different conceptions of the good might live together. Let us recall Rawls’s definition of a “comprehensive view,” according to which “it includes conceptions of what is of value in human life, as well as ideals of personal virtue and character, that are to inform much of our nonpolitical conduct (in the limit our life as a whole)” (*PL*, p. 175). On this criterion, *A Theory of Justice* is already “political” as against “comprehensive.”

III. THE PROBLEM OF STABILITY

Rawls’s own account of what was wrong with *A Theory of Justice* does not stand up. Nevertheless, I accept his claim that it is a concern for stability that has impelled him along his new course. In *Political Liberalism*, he says that “the problem of stability has played very little role in the history of moral philosophy” even though it “is fundamental to political philosophy” (*PL*, p. xvii). Perhaps it is true that the actual word “stability” has not been used commonly in the history of political philosophy, though the *Oxford English Dictionary* traces uses of it in a social and political context back in the fifteenth century, citing Malory as writing of “the stabylyte of this realme.”⁹ But all we have to do is rechristen the problem of stability as the problem of order and we can immediately recognize it as a central focus of political philosophy in all periods—but especially, of course, at times when order is particularly problematic. Hobbes, reacting to the English Civil War, and de Maistre, reacting to the French Revolution, might be seen as the archetypal theorists of the problem of order. This suggests that the problem of order is the property of the political right, with the solution lying in the unfettered exercise of despotic power. (It was, indeed, a commonplace in the nineteenth century to think of the opposing political forces as those of “progress” and “order.”) But there is a more egalitarian strand, running through Rousseau and Durkheim, according to which the solution of the problem lies in some sort of normative consensus among the members of a society. Rawls manifestly belongs to this school rather than that of Hobbes and de Maistre.

In his first book, *The Structure of Social Action*, Talcott Parsons took engagement with the Hobbesian problem of order as definitive of

8. *Ibid.*, referring to *A Theory of Justice*, p. 17.

9. Sir Thomas Malory, *Le morte d'Arthur* (trans. 1470–85), cited in the *Oxford English Dictionary*, 2d ed., s. v. “stability,” 2a. Political scientists have been concerned for many years with the stability of regimes, using that word to identify the problem.

sociology as a discipline. This problem is constituted by the fact that people in a Hobbesian state of nature have (in Parsons's terminology) "random ends."¹⁰ The question that this poses is how to arrange matters so that these different ends do not lead people into conflict with one another. We can readily translate this into Rawls's terminology: given, we may say, a plurality of conceptions of the good (or "comprehensive views"), how can the members of a society be motivated so as to have a "strong and normally effective desire" to act in ways demanded by their common institutions?¹¹ Formally, Rawls's solution is the same as Hobbes's: that people should retain their differing ends (or conceptions of the good) but reach agreement on certain ideas about what justice requires. Where he departs from Hobbes is in the motivation he seeks for adhering to the dictates of justice.

I take Hobbes to rely primarily on a general sense of the preferability of civil order (any civil order) to a relapse into a "state of nature." Rawls, however, makes it clear that he is not interested in the kind of stability that may arise when supporters of conflicting conceptions of the good reach a standoff and settle for Rawlsian justice as the best they can hope for: "the conception of justice as fairness is not regarded merely as a *modus vivendi*."¹² Nor is it a matter of finding "means of persuasion or enforcement."¹³ Why does Rawls renounce the quest for stability arising in these ways? One possible answer (which was in essence the one given by Parsons) is that the motive of fear cannot be relied on to secure peace. This is not Rawls's answer. He does not, as far as I am aware, ever deny that stability might be secured by Hobbesian means. Even if it could, though, this would be irrelevant from his point of view. There are two reasons for this.

The first reason is that Rawls is concerned not with the stability of societies in general but with the stability of just societies. Since a just society has civil liberty and democratic participation in government, it cannot be maintained by Hobbesian methods. Thus, Rawls states as one of the "general facts" to be taken into account in formulating a theory of justice "that an enduring and secure democratic regime, one

10. Talcott Parsons, *The Structure of Social Action* (Glencoe, Ill.: Free Press, 1937), see esp. "Hobbes and the Problem of Order," pp. 89–94. For a discussion of Parsons on the "problem of order," see my *Sociologists, Economists, and Democracy* (reprint; Chicago: University of Chicago Press, 1978), pp. 76–86.

11. The quoted expression is drawn from Rawls's standard definition of the "sense of justice": it is "an effective desire to apply and to act from the principles of justice and so from the point of view of justice" (*TJ*, p. 567). I shall discuss it in the next section of this article.

12. John Rawls, "Justice as Fairness: Political not Metaphysical," *Philosophy and Public Affairs* 14 (1985): 223–51, 247; see also *PL* pp. 146–49. Hereafter, cited as "Justice as Fairness" (Rawls's earlier article with that title is not discussed here).

13. Rawls, "The Domain of the Political," p. 246; see also *PL*, p. 142.

not divided into contending doctrinal confessions and hostile social classes, must be willingly and freely supported by at least a substantial majority of its politically active citizens" (*PL*, p. 38). As we shall see, Rawls has always accepted (and still does) that some people may simply have to be coerced into acting as justice demands. But just institutions cannot be maintained if these people form more than a small minority of the population.¹⁴

It can be said that this is still a pragmatic rather than a principled reason for rejecting Hobbesian order as a solution to the problem of stability. But Rawls's second objection is one of fundamental principle. "Reasonable persons," Rawls tells us in *Political Liberalism*, "desire for its own sake a social world in which they, as free and equal, can cooperate with others on terms all can accept" (*PL*, p. 50). This is not a new claim on Rawls's part. In *A Theory of Justice*, he frequently emphasizes that his objective is the practical one of proposing terms for living together that can gain the assent of all the members of a society. Thus, he says that "it is partly" in order to reach "the greatest convergence of opinion" that "we accept the constraints of a common standpoint, since we cannot reasonably expect our views to fall into line when they are affected by the contingencies of our different circumstances" (*TJ*, p. 517). He goes on to emphasize that "the numerous simplifications of justice as fairness" (*TJ*, p. 517) are justified by their role in making agreement more feasible. The hope of the parties in the original position is that, by settling for two relatively simple principles lexicographically ordered, they will succeed in "simplify[ing] political and social questions so that the resulting balance of justice, made possible by the greater consensus, outweighs what may have been lost by ignoring certain potentially relevant aspects of moral situations" (*TJ*, p. 517). Thus, the requirement that the principles of justice should form the publicly acknowledged charter of a society sets limits on how complex they can be.¹⁵

For Rawls, then, stability is not merely a matter of general compliance with the rules; it has to be compliance for the right reasons. By definition, only a society in which the terms of cooperation embodied in the rules are accepted freely can count as a stable society. In a stable society, as Rawls defines it, "citizens act willingly so as to give one another justice over time. Stability is secured by sufficient motivation

14. The point has recently been made in the following terms: "A differentiated society cannot work if elementary freedoms of movement and association for all are to be preserved. . . . The disaffected cannot overthrow society, but they can make it impossible to live in" (Paul Hirst, *Associative Democracy: New Forms of Economic and Social Governance* [Oxford: Polity, 1994], pp. 9–10).

15. Most of this paragraph is drawn from p. 166 of my *Justice as Impartiality* (Oxford: Clarendon, 1995), since I cannot think of any way of putting it better.

of the appropriate kind acquired under just institutions” (*PL*, pp. 142–43).

IV. THE GOOD OF JUSTICE

If this is what stability is, how is it to be secured? In *A Theory of Justice*, Rawls gives two answers, one in chapter 8 and one in chapter 9 (the last chapter in the book). Chapter 8 is on the development of the “sense of justice,” a “strong and normally effective desire to act as the principles of justice require” (*TJ*, p. 454). The standard account of Rawls on motivation would, I think, run along some such lines as these: people (or at any rate well brought up people in a just society) have a “sense of justice,” which is a disposition to behave justly; and, since they accept a theory in which the right has ethical priority over the good, their sense of justice leads them to give the demands of justice priority over the pursuit of their good, when the two come into competition. This account draws on chapter 8 and ignores chapter 9. If we want to present Rawls’s theory as an attractive one, there is much to be said for sticking to it. But if we want to trace the path that led Rawls to “overlapping consensus,” we have to focus on chapter 9. This, as we shall see, requires us to revise all the features of the standard account just given.

To begin with, we have to be clear that “the priority of the right over the good” is a theoretical proposition, forming part of the architecture of the theory of justice. It has nothing to say about the strength of the motivation to comply with the demands of justice. In particular, it does not say that the recognition of something as required by justice will provide an overriding motive for doing it.¹⁶ For that would invoke the “doctrine of the pure conscientious act,” associated with Sir David Ross and H. A. Prichard, which Rawls says “we should not rely on” (*TJ*, p. 569).

Now, the most distinctive point of Prichard’s famous article “Does Moral Philosophy Rest on a Mistake?” was to deny that right acts have anything in common that makes them right.¹⁷ In particular, there is no end served by all right actions. This is, as far as it goes, Rawls’s

16. Michael Sandel’s critique in *Liberalism and the Limits of Justice* (Cambridge: Cambridge University Press, 1982) trades on attributing to Rawls an idea—“the primacy of justice”—that he does not employ and treating it as a portmanteau of the “priority of right” and “congruence.” A recent restatement is even more blatant: here, Sandel silently changes “the priority of the right over the good” (an ethical priority) to “the priority of the right to the good” (a temporal priority, if it makes any sense at all), and then equates this with the motivational claim that “considerations of justice will always outweigh other, more particular ones” (Michael J. Sandel, “Political Liberalism,” *Harvard Law Review* 107 [1994]: 1765–94, p. 1769).

17. Originally published in 1912, collected in *Moral Obligation: Essays and Lectures*, ed. W. D. Ross (Oxford: Clarendon, 1949).

own doctrine of the priority of the right over the good, which he insisted on in *A Theory of Justice* and still adheres to in *Political Liberalism*. Thus, he says in the latter that “the priority of right means . . . that the principles of justice set limits to permissible ways of life” and concludes from this that “the claims that citizens make to pursue ends transgressing those limits have no weight” (*PL*, p. 209).

What Rawls is focusing on, however, is not the Prichard/Ross antiteleological conception of the right (which he shares) but their ideas about the motivation for right action. These are not actually central to their writings.¹⁸ In brief, they held that recognizing something to be right is sufficient to motivate right action. Rawls treats this as if it made acting rightly a kind of *acte gratuit*. Thus, he says that the “doctrine of the purely conscientious act is irrational” (*TJ*, p. 477). For “on this interpretation the sense of right lacks any apparent reason; it resembles a preference for tea rather than coffee” (*TJ*, p. 478).

I am inclined to think that this is a travesty of the thoroughly commonsensical idea represented by saying that people can do their duty out of a sense of duty and not in order to achieve some independently definable end. This does not have to mean that moral action is “utterly capricious,” as Rawls has it (*TJ*, p. 478). We can perfectly well tell a story about motivation that makes acting rightly appear as rational. The story that I would especially commend is that of T. M. Scanlon. According to this, the moral motive is the desire to act according to rules that could not reasonably be rejected by others similarly motivated.¹⁹ Since this desire is widespread, we are predisposed to act rightly. But this predisposition does not give rise to any action until we have determined what actually is the right thing to do. It is therefore quite natural to say that the thought that something is the right thing to do is what motivates us to act rightly. I take this to be precisely the proposition that Rawls objects to. But it is quite innocuous against the background of a standing desire to act in the ways we think it right to act.

We can now see the rationale for a chapter that has the title “The Good of Justice.” In both the old and the new forms of Rawls’s theory, the “priority of the right over the good” is asserted, but that assertion is taken to leave us with the question of why people might actually

18. Neither the entry for Prichard nor the entry for Ross in the *Encyclopedia of Ethics* mentions their ideas about moral motivation, talking instead about their antiteleological conception of ethics and their intuitionistic explanation of the way we come to know that acts of certain kinds are right. See *Encyclopedia of Ethics*, ed. Lawrence C. Becker and Charlotte B. Becker (New York: Garland, 1992), vol. 2, s. v. “Prichard,” “Ross.”

19. See T. M. Scanlon, “Contractualism and Utilitarianism,” in *Utilitarianism and Beyond*, ed. Amartya Sen and Bernard Williams (Cambridge: Cambridge University Press, 1982), pp. 103–28.

give the right priority in practice. Recoiling from “the doctrine of the pure conscientious act,” Rawls commits himself in chapter 9 of *A Theory of Justice* to the ancient doctrine that no act can be regarded as rational unless it is for the good of the agent to perform it. Thus, the problem is one of “congruence” between justice and goodness. Each of “the concepts of justice and goodness . . . with its associated principles defines a point of view from which institutions, actions, and plans of life can be assessed” (*TJ*, p. 567). Rawls adds, significantly, that “whether these two points of view are congruent is likely to be a crucial factor in determining stability. But congruence is not a foregone conclusion even in a well-ordered society” (*TJ*, p. 567).

Rawls presupposes that the theory of justice is itself established: the principles of justice would rationally be chosen in the original position and hence we can say that “just institutions are collectively rational and to everyone’s advantage from a suitably general perspective” (*TJ*, p. 567). Within his current terminology, which makes a sharp distinction between rationality and reasonableness, Rawls would presumably wish to say that the rationality of choosing the principles of justice in an appropriately constituted original position suggests that they form a reasonable basis for agreement among real people. Either way, the point is that he assumes a gap to exist between accepting the principles of justice and being motivated to act on them. Even if people have a “sense of justice,” it still has to be “shown that this disposition to take up and to be guided by the standpoint of justice accords with the individual’s good” (*TJ*, p. 567). Thus, the problem of congruence can be formulated (as in the title of the penultimate section of *A Theory of Justice*) by saying that it is a question of “the good of the sense of justice.”

V. THE CONDITIONS OF CONGRUENCE

Rawls makes a false start in this penultimate section by describing “the real problem of congruence” as “what happens if we imagine someone to give weight to his sense of justice only to the extent that it satisfies other descriptions which connect it with reasons specified by the thin theory of the good” (*TJ*, p. 569). If Rawls were to follow through this line of analysis, the task before him would be straightforward, since the “thin theory” (if it works at all) applies to everyone. As he himself says, if “the regulative desire to adopt the standpoint of justice” is “rational for one, it is rational for all, and therefore no tendencies to instability exist” (*TJ*, p. 567). Rawls adds, however, that the search for reasons is to be carried on “in the light of the thin theory with no restrictions on information” (*TJ*, p. 567). This would seem to me to make the “thin theory” irrelevant. Given that the problem, as Rawls conceives it, is one of offering a reason for real people to “affirm their sense of justice,” it is surely correct to specify that they should have

full information. But then why should they throw away the information about their own distinctive conceptions of the good (their “thick” conception) and restrict the question to one the answer to which is going to be the same for everyone?

Rawls does not explain. But in the event nothing turns on it. For the thin theory of the good in the context of full information apparently amounts to no more than means-end rationality. This, as Rawls goes on later to concede, leaves it open to us to “suppose that even in a well-ordered society there are some persons for whom the affirmation of their sense of justice is not a good. Given their aims and wants and the peculiarities of their nature, the thin account of the good does not define reasons sufficient for them to maintain this regulative sentiment” (*TJ*, p. 575). This is manifestly abandoning the idea that appealing to the “thin theory of the good” means that if “the regulative desire to adopt the standpoint of justice” is “rational for one, it is rational for all” (*TJ*, p. 567).

What are the reasons that might lead us to hope for congruence to obtain between the requirements of justice and the good of at any rate the great majority of people in contemporary society? Rawls rehearses the argument from much earlier in the book about the relative “strains of commitment” generated by his principles of justice and the principle of maximizing utility (*TJ*, pp. 175–83). But he puts it in a remarkably downbeat form, saying merely that “however improbable the congruence of the right and the good in justice as fairness, it is surely more probable than on the utilitarian view” (*TJ*, p. 573). The case for congruence is strengthened by adducing the psychological cost of dissimulation in a well-ordered society and the risk of hurting those to whom we are in some way attached if we behave unjustly. Rawls adds that “to share fully in” the life of our society, and enjoy “the greater richness and diversity” made possible by cooperation, “we must acknowledge the principles of its regulative conception, and this means that we must affirm our sentiment of justice” (*TJ*, p. 571).

Having set out what he calls “the chief reasons . . . for maintaining one’s sense of justice,” Rawls says that “the question now arises whether they are decisive” (*TJ*, p. 572). And his answer is in the negative: all that can be said is that “we confront the familiar difficulty of a balance of motives” (*TJ*, p. 572). There is, however, one way in which decisiveness can be guaranteed, and that is via the “Kantian interpretation” of the theory of justice. According to this, “acting justly is something we want to do as free and equal rational beings” (*TJ*, p. 572). The problem of decisiveness is then solved definitively, since we can express our nature as free and equal rational beings only to the extent that we commit ourselves unconditionally to give our sense of justice precedence over our other interests. “This sentiment cannot be fulfilled if

it is compromised and balanced against other ends as but one desire among the rest. It is a desire to conduct oneself in a certain way above all else, a striving that contains within itself its own priority" (*TJ*, p. 574; see also p. 476, where Rawls says that for people to "express their nature as free and equal rational beings . . . belongs to their good.")

VI. STABILITY AND LEGITIMACY

I have argued, against Rawls, that neither the theory of justice nor the conception of a "well-ordered society" in *A Theory of Justice* depends upon the acceptance of a (fully or partly) "comprehensive view," as defined by him. Suppose, however, that the stability of a just society can be securely underpinned only by appealing to the "Kantian interpretation" of the principles of justice. This must count as a "comprehensive view" inasmuch as it invokes a particular conception of the good: a conception according to which the highest good of human beings is to express their free and equal rational natures, which they can do only by committing themselves in advance to acting justly. Hence, we could say that, although the theory of justice itself does not depend upon a "comprehensive view," the account of its stability does.²⁰

It is not hard to see how further reflection over the years might have led Rawls to feel dissatisfied with such an account. For the requirement of "congruence" is an extraordinarily demanding one. It can be met unconditionally by invoking the "Kantian interpretation." But even if that contains the truth about human beings (an assertion to which Rawls never commits himself unequivocally in *A Theory of Justice*), it is clear that he does not regard it as a straightforward solution to the problem of congruence. If it were, there could not possibly be "many who do not find a sense of justice for their good," as Rawls says there could be (*TJ*, p. 576). Perhaps Rawls thinks that the "Kantian interpretation" is a motivating force only among those who believe in

20. Samuel Freeman arrives at the same conclusion, writing that "from a reading of the first five hundred pages (chapters 1–8) of *A Theory of Justice*, it is hard to see exactly where Rawls thinks he had to invoke a more comprehensive ethical doctrine to justify justice as fairness." It is, he says, "only when we reach chapter 9 and the second stage of the argument for stability, from 'congruence,' that the deeper bases of the view in Kantian ethics become really apparent" (Samuel Freeman, "Political Liberalism and the Possibility of a Just Democratic Constitution," *Chicago-Kent Law Review* 69 [1994]: 619–68, 628). The aptness of these remarks illustrates Sandel's perversity in reading *A Theory of Justice* as if the argument for the principles of justice depended on the "Kantian interpretation." This is the inevitable consequence of his treating the "priority of justice" (which is part of the theory itself) as a proposition about motivation, thus spreading the Kantian infection back through the whole book from chap. 9. Ironically, the terms of Rawls's *mea culpa* (discussed above in Sec. II) suggest that Sandel may have succeeded in converting him to this misreading of *A Theory of Justice*!

it. This would be equivalent to his current notion (see Sec. XII) that a Kantian “comprehensive view” will underwrite Rawlsian “justice as fairness,” but that we cannot expect everybody to subscribe to this view.

If I am right, Rawls can hardly be said to have solved the problem of stability in *A Theory of Justice* as he defines it there. This piece of unfinished business might by itself have been enough to impel Rawls into taking another shot at the problem of stability. But I believe that it is possible to offer a further explanation. For Rawls seems to have changed his mind in one respect about the significance of stability. I have quoted him as saying that we are to seek stability of a non-Hobbesian kind. But what if non-Hobbesian stability cannot be attained? In *A Theory of Justice*, Rawls has no qualms about falling back on Hobbesian methods. Thus, the penultimate section, “The Good of the Sense of Justice,” concludes with a robust insistence that

to justify a conception of justice we do not have to contend that everyone, whatever his capacities and desires, has a sufficient reason (as defined by the thin theory) to preserve his sense of justice. For our good depends upon the sort of persons we are, the kinds of wants and aspirations we have and are capable of. It can even happen that there are many who do not find a sense of justice for their good; but if so, the forces making for stability are weaker. Under such conditions penal devices will play a much larger role in the social system. The greater the lack of congruence, the greater the likelihood, other things equal, of instability with its attendant evils. [TJ, p. 576]

Given his definition of stability at the beginning of chapter 9, Rawls should not really be talking here about lack of congruence having a *tendency* to produce instability: he should, to be consistent, say that it *constitutes* it. But the point that I want to focus on here is that coercion raises no ethical problem in *A Theory of Justice*. To see why Rawls sees only practical problems in coercing those who do not find justice for their good, let us go back to the first chapter. Here he says that, in accepting the principles of justice, people agree at the same time “to conform their conceptions of the good to what the principles of justice require, or at least not to press claims which directly violate them. . . . The principles of right, and so of justice, put limits on which satisfactions have value; they impose restrictions on what are reasonable conceptions of one’s good. . . . A just social system defines the scope within which individuals must develop their aims, and it provides a framework of rights and opportunities and the means of satisfaction within and by the use of which these ends may be equitably pursued” (TJ, p. 31).

It follows from this that people whose good is inherently incompatible with justice can be accused of having cultivated (or not having worked to extinguish) an unreasonable conception of the good. Rawls

gives the example of somebody “who finds that he enjoys seeing others in positions of lesser liberty.” Since he must “accept in advance a principle of equal liberty” (on the basis of the arguments for justice as fairness), he “understands that he has no claim whatever to this enjoyment. The pleasure he takes in other’s [*sic*] deprivations is wrong in itself: it is a satisfaction which requires the violation of a principle to which he would agree in the original position” (*TJ*, p. 31). Someone who has an unreasonable conception of the good in this sense cannot legitimately object if his attempts to pursue it are suppressed by force, since its successful pursuit must by its nature lead to the perpetration of injustice.

Inherently unjust conceptions of the good are important but far less common than conceptions of the good whose pursuit beyond some point is contingently unjust. Thus, Rawls’s “primary goods” are things that, according to him, it is rational to want more rather than less of. It is, I suggest, perfectly reasonable for us to want primary goods beyond what is just—so long as we do not want them under a description that includes the amount’s being unjust. We can accept that justice sets limits on the pursuit of something we conceive as being for our good, but we do not have to abandon the view that it is for our good.

Suppose that I form the view that it would contribute to my good to take a trip around the world, and that I then find that this would cost more than my resources permit. (Let us follow Rawls in assuming that I live in a just society, so my budget limit corresponds to the one imposed by just economic institutions.) Instead of simply concluding that I cannot justly take the trip (while continuing to believe that taking it would be for my good), I am told by Rawls that I must somehow persuade myself that it would not be for my good at all. For only that thought can motivate me to refrain from taking the trip unjustly if the opportunity should arise. This is the absurdity into which Rawls is led by his rejection of “the doctrine of the purely conscientious act.”²¹

The upshot is, then, that only conceptions of the good that are congruent with justice are reasonable. Anyone who holds any other conception of the good can legitimately be coerced—and will have to be. Nobody can reasonably object to this because everyone must recognize that “it is rational to authorize the measures needed to maintain just institutions, assuming that the constraints of equal liberty and the rule of law are duly recognized” (*TJ*, p. 576).

21. Rawls makes fleeting contact with common sense when he says that, as an alternative to agreeing “to conform their conceptions of their good to what the principles of justice require,” people may agree “at least not to press claims which directly violate them” (*TJ*, p. 31). But this idea is never followed up. We should not be surprised, because it conflicts with Rawls’s doctrine that nobody has any reason for acting in ways contrary to his own good.

As we shall see (in Sec. VIII), Rawls does not now think that “comprehensive views” incompatible with justice are ipso facto unreasonable. And with this change of mind goes a correlative one about the significance of stability, understood as congruence or as “overlapping consensus.” We are now told that “justice as fairness is not reasonable in the first place unless in a suitable way it can win its support by addressing each citizen’s reason, as explained within its own framework” (*PL*, p. 143). Thus, stability and legitimacy are now tied together in a new way that raises the stakes a lot. In *A Theory of Justice*, the validity of the theory was unaffected by any lack of success it might have in forming a part of people’s conceptions of their good. For we could say that all conceptions of the good incompatible with the principles of justice were necessarily unreasonable. But in *Political Liberalism*, Rawls denies that a regime can be legitimate if there exist people with “reasonable comprehensive views” who reject its foundational principles.

I said earlier that dissatisfaction with his solution to the problem as stated in chapter 9 of *A Theory of Justice* might have been enough to lead Rawls into putting forward a revised account of stability as “congruence.” I can now add that Rawls’s change of mind about the relation between stability and legitimacy makes the importance of succeeding much greater. In *A Theory of Justice*, lack of “congruence” would be an inconvenience, because those with unreasonable views (conceptions of the good incompatible with the demands of justice) might have to be coerced. In *Political Liberalism*, it would be a catastrophe if there were people with “reasonable comprehensive views” that could not accommodate the demands of justice, since that would undermine the legitimacy of Rawlsian institutions.

VII. THE “POLITICAL CONCEPTION” OF JUSTICE

We have seen that, in the last chapter of *A Theory of Justice*, Rawls maintained that the stability of the principles of justice could be assured only to the extent that they were “congruent” with widely held conceptions of the good. In *Political Liberalism*, he says that the stability of the principles of justice can be assured only to the extent that they are in some way compatible with widely held “comprehensive views.” Apart from a shift in terminology, it is clear that in this respect Rawls has not changed his mind about stability. (If we stipulate that one element of “communitarianism” is skepticism about the motivational force of Rawlsian-style justice when it conflicts with conceptions of the good, we can say that Rawls has always been a “communitarian.”)²²

22. A more distinctive “communitarian” tenet might be taken to be that conceptions of the good are not subject to appraisal or revision but are simply part of people’s identity. In this sense Rawls never has been a “communitarian” and is not one now.

Nevertheless, it is the ideas about stability that differentiate *Political Liberalism* from *A Theory of Justice*. Thus, Rawls says that, apart from the new account of stability, “these lectures take the structure and content of *Theory* to remain substantially the same” (*PL*, p. xvi).

In the introduction to *Political Liberalism*, Rawls says that he was puzzled by objections to his initial statements of the ideas of “a political conception of justice and of an overlapping consensus.” He tells us that he now thinks that this was because he “had taken for granted a few missing pieces,” the first of which is “the idea of justice as fairness as a freestanding view and that of an overlapping consensus as belonging to its account of stability” (*PL*, p. xxx). I do not believe that the objections to “overlapping consensus” are surmounted by emphasizing that it is an account of the stability of justice rather than an account of what justice is. Nevertheless, it obviously helps to put the core ideas of *Political Liberalism* into perspective to acknowledge that they are intended to replace the account of stability in *A Theory of Justice* and not the whole book.

We may thus follow Rawls in thinking of the argument to be made in favor of a conception of justice as having two stages. “In the first stage,” he says, justice as fairness “is worked out as a freestanding political (but of course moral) conception for the basic structure of society” (*PL*, pp. 140–41). To say that it is “a political conception” is simply to make the point that it does not purport to cover the whole of life. (See Sec. II above.) Its principles are concerned with the basic structure of society: its major political, economic, and social institutions. Within that framework, it is to be assumed that different conceptions of the good will lead people to pursue widely diverse objectives, either alone or in association with others. The second stage concerns the stability of the principles of justice arrived at in the first stage. Although my focus here is (as is Rawls’s in *Political Liberalism*) on the second stage, I shall be able to make my argument about the second stage (namely, that we don’t need it) only by saying something about what goes into the first stage.

Although Rawls emphasizes the importance of distinguishing the two stages in the arguments for “justice as fairness,” he himself encourages just the kind of misunderstanding he denounces, by failing to provide anywhere in *Political Liberalism* a coherent and self-contained statement of what he takes the case for “justice as fairness as a freestanding view” to be. A few years earlier, in “The Domain of the Political and Overlapping Consensus,” a quite comprehensive statement of the themes of *Political Liberalism*, Rawls did set out the two stages sequentially. It is rather sad to see how, in *Political Liberalism*, this article is dismembered, with paragraphs from the exposition of the first stage scattered around the book and almost invariably mixed up with material relevant to the second stage. The effect of this is,

perversely, to undermine the integrity of the first stage that Rawls says he wishes to emphasize. Since the article is truer to Rawls's stated intentions and does not appear to contain anything inconsistent with the doctrines of *Political Liberalism*, I shall follow its exposition of the first stage, supplying cross-references as appropriate.

We begin with four "general facts" that, according to Rawls, must be taken into account in constructing a conception of justice appropriate to a contemporary society. The first of these is "the fact of pluralism"—the "diversity of conflicting and irreconcilable comprehensive doctrines" that is bound to exist so long as a society has "free institutions." The "second and related general fact is that only the oppressive use of state power can maintain a continuing common affirmation of one comprehensive religious, philosophical, or moral doctrine." The third is that "an enduring and secure democratic regime . . . must be willingly and freely supported by at least a substantial majority of its politically active citizens." The fourth fact is that "the political culture of a reasonably stable democratic society normally contains, at least implicitly, certain fundamental intuitive ideas from which it is possible to work up a political conception of justice suitable for a constitutional regime."²³ Rawls adds, after some further discussion, "a fifth general fact: we make many of our most important judgments subject to conditions which render it extremely unlikely that conscientious and fully reasonable persons, even after free discussion, can exercise their powers of reason so that all arrive at the same conclusion."²⁴

What we have to look for, then, is a conception of justice capable of being accepted by people with a diversity of (religious or secular) views which should build on "fundamental intuitive ideas" such as "the idea of society as a fair system of social cooperation over time from one generation to the next, and the idea of citizens as free and equal persons fully capable of engaging in social cooperation over a complete life."²⁵ The answer is, we already know, Rawls's principles of justice. But how do we get to them?

The argument proceeds, as in *A Theory of Justice*, through the device of an "original position," in which the principles of justice are chosen by agents from behind a "veil of ignorance." In "The Domain of the Political," this is introduced without any explanation or justifica-

23. Rawls, "The Domain of the Political," pp. 234–35. In *Political Liberalism* the first three "general facts" are set out on pp. 36–38 and the fourth in n. 41 on p. 38, but (tendentiously) in a way that connects them up with the argument for "overlapping consensus."

24. Rawls, "The Domain of the Political," p. 238 (repeated, with slight verbal changes, on p. 58 of *Political Liberalism*).

25. *Ibid.*, p. 240. The first point occurs on pp. 15 and 18 of *Political Liberalism*; the second is diffused over pp. 16–19.

tion, but in *Political Liberalism* we get a quite extensive rationale (*PL*, pp. 22–28). “The original position is simply a device of representation: it describes the parties, each of whom is responsible for the essential interests of a free and equal citizen, as fairly situated and as reaching an agreement subject to conditions that appropriately limit what they can put forward as good reasons” (*PL*, p. 25).²⁶ The point of the “veil of ignorance” is to exclude from deliberation those features that should be regarded as irrelevant to it. Thus,

the fact that we occupy a particular social position is not a good reason for us to propose, or to expect others to accept, a conception of justice that favors those in this position. Similarly, the fact that we affirm a particular religious, philosophical, or moral comprehensive doctrine with its associated conception of the good is not a reason for us to propose, or to expect others to accept, a conception of justice that favors those of that persuasion. To model this conviction in the original position, the parties are not allowed to know the social position of those they represent, or the particular comprehensive doctrine of the person each represents. [*PL*, p. 24]

Having got this far, Rawls leaves us in the lurch in *Political Liberalism*. Early on in the book, he restates the principles of justice arrived at in *A Theory of Justice* (*PL*, pp. 5–6), and says that “all these elements [of his egalitarian form of liberalism] are still in place, as they were in *Theory*; and so is the basis of the argument for them” (*PL*, p. 7). We can only assume that this argument is the one to be found in *A Theory of Justice*. In “The Domain of the Political,” however, we are offered a sketch of an argument for the first principle of justice. Thus, if we “suppose in the first stage that the parties [in the original position] assume the fact of pluralism to obtain . . . [they] must then protect against the possibility that the person each party represents may be a member of a religious, ethnic, or other minority. This suffices for the argument for the equal basic liberties to get going.”²⁷

Rawls does not explain how the second principle (the one that deals with social and economic inequalities) is to be derived, but the same basic motivation can be invoked to support both principles. As Rawls explains, the people in the original position endorse equal rights because they are more interested in avoiding the risk of being discriminated against as a minority than in the prospect of being able to oppress others as a member of a majority. Similarly, we may assume that the same concern for avoiding the worst will lead to their being

26. Anyone interested in Rawlsian stratigraphy may care to note that this comes from p. 237 of “Justice as Fairness.”

27. Rawls, “The Domain of the Political,” p. 251.

anxious to rule out in advance the possibility of being reduced to a level of poverty that nobody need endure. And this concern will be met by their endorsing the difference principle, which calls for social institutions to be arranged for the benefit of the worst off group.

I have argued elsewhere that Rawls actually weakens the argument for his principles of justice by forcing it through the machinery of the original position.²⁸ For he can get the conclusions he wants out of it only by postulating extreme risk aversion among the agents doing the choosing, and this has an inevitably gimmicky air about it. As one of Rawls's students, Peter de Marneffe, has correctly said, this extreme risk aversion is really no more than a way of representing the idea that inequalities have to be justified to everybody, and especially to those who come out at the bottom.²⁹ But this argument is better made directly, omitting entirely the contortions necessary to fit it into the choice of principles from behind a "veil of ignorance."

If we go back to Rawls's explanation in *Political Liberalism* of the distinctive features of his "original position," we might say that he offers us something more useful than an attempt to extract the principles of justice from it—a suggestion that the moral premises required to justify its taking the form it does are sufficient to enable us to arrive at the principles of justice without it. To see this, let us pick up again Rawls's claim that the original position is a device for representing good reasons. It is, Rawls says, not a reason we can expect others to accept that some principle of justice favors those in our position (including in this both natural and social advantages) or those who hold our own particular "religious, philosophical or moral comprehensive doctrine with its associated conception of the good" (*PL*, p. 24). If we start from these premises, we can move directly to an argument for Rawls's two principles of justice.

Thus, it may be argued that the only terms that cannot reasonably be rejected for living together by the adherents of any "comprehensive view" are those embodied in Rawls's first principle of justice. For only "equal freedom" is neutral in an appropriate way between different conceptions of the good. Hence, nobody can claim to be unfairly discriminated against by it. The second principle of justice flows from the premise that no valid claim can be made on the basis of natural or social advantages. For then the only reason that can be advanced for departing from an equal distribution of socioeconomic primary goods is that this is in the interests of all. People cannot be expected

28. See my *Theories of Justice*, pp. 213–15, and *Justice as Impartiality*, pp. 52–61.

29. Peter de Marneffe, "Liberalism, Liberty and Neutrality," *Philosophy and Public Affairs* 19 (1990): 253–74.

to accept an inequality that leaves them at the bottom unless they can be shown that it works to the advantage of the worst off.³⁰

This is, obviously, the merest outline of an argument. I hope, however, that it is enough to give substance to my claim that we can get from Rawls's premises to his conclusions without the suspect intermediation of his "original position." It may not have escaped notice that the form taken by the argument is foreshadowed in the version of contractarian theory put forward by T. M. Scanlon, which "takes the fundamental question to be whether a principle could reasonably be rejected . . . by parties who, in addition to their own personal aims, were moved by a desire to find principles that others similarly motivated could also accept."³¹ The convergence is not accidental because Scanlon's approach may be seen as what we are naturally led to if we are attracted by Rawls's general approach but regard his "original position" as an embarrassment.

In Rawls's brief exposition, we start from the "fact of pluralism." On the basis of this, we make the following move: "Since many doctrines are seen to be reasonable, those who insist, when fundamental political questions are at stake, on what they take as true but others do not" are making "a claim that cannot be made good by anyone to citizens generally" (*PL*, p. 61). We then argue to the conclusion that people seeking agreement on reasonable terms with others must concede that the pursuit of their comprehensive views must be constrained by the limits laid down by Rawlsian justice. This conclusion will carry weight, of course, only with those who are seeking agreement on reasonable terms with others. But the first stage includes in it the resources for underwriting this motivation.

I have already quoted Scanlon on the "moral motive" as the desire to live according to rules that could not reasonably be rejected by others. Rawls says in *Political Liberalism* that "in setting out justice as fairness we rely on the kind of motivation Scanlon takes as basic" (*PL*, p. 50, n. 2). And he devotes a section (*PL*, pp. 81–86) to "The Basis of Moral Motivation in the Person." This serves the same function as chapter 8 of *A Theory of Justice* inasmuch as it concerns itself with the "sense of justice": the disposition to act as justice requires. "Thus," he concludes the section, "besides a capacity for a conception of the good,

30. This is not a prudential but a moral argument. The point is not that *these particular* worst off people could not be made better off but that the worst off (whoever they may be) are as well off as possible. (Rawls makes this point in the course of his exposition of the principles of justice: *PL*, p. 7, n. 5.)

31. T. M. Scanlon, "Levels of Moral Thinking," in *Hare and Critics*, ed. Douglas Seanor and N. Fotion (Oxford: Clarendon, 1982), pp. 103–28, pp. 137–38.

citizens have a capacity to acquire conceptions of justice and fairness and a desire to act as these conceptions require" (*PL*, p. 86).

It must be emphasized that Scanlonian motivation is invoked as an element in the "political conception." It is thus taken to be established before we get to the search for stability that makes up the second stage. But why, in that case, do we need the second stage? If the problem of stability is one of showing that reasonable people have adequate motives for adhering to the demands of Rawlsian justice, this would appear to have been secured by the first-stage argument. To see why Rawls thinks the search for stability still needs to go on, we need to examine his account of the second stage. That is the task to which I turn next.

VIII. "REASONABLE COMPREHENSIVE VIEWS"

The stability of "justice as fairness" can be guaranteed, Rawls claims, only if it falls within an "overlapping consensus" of "reasonable comprehensive views." This could easily be accomplished by stipulating that only "comprehensive views" compatible with Rawls's principles of justice were to be regarded as reasonable.³² Occasionally, Rawls says things that would appear to make the reasonableness of a view depend on its content in this way. Thus, at one point in *Political Liberalism*, Rawls makes the usual first-stage argument that "when equally represented [in the original position], no citizen could grant to another person or association" a form of authority that would give it "the right to use state power to decide constitutional essentials as that person's, or that association's, comprehensive doctrine directs." But he adds that "any such authority is, therefore, without grounds in public reason, and reasonable comprehensive doctrines recognize this" (*PL*, p. 226). Elsewhere, he says that "a comprehensive doctrine is not as such unreasonable because it leads to an unreasonable conclusion in one or even in several cases. It may still be reasonable most of the time" (*PL*, p. 244, n. 32). This introduces a new idea (not pursued anywhere else in *Political Liberalism*), that doctrines may be more or less reasonable, rather than simply reasonable or unreasonable. The point is, however, that it too provides a content-based criterion for a reasonable doctrine, since it appears that a doctrine is to count as reasonable only to the extent that the practical conclusions drawn from it square with those to be drawn from Rawls's principles of justice.

To set against these isolated passages, we have a number of reasons for thinking that Rawls wants to leave it as a conceptually open

32. I have observed in discussion that this idea about *Political Liberalism* is quite common. It can be found in Susan Moller Okin, "Political Liberalism, Justice, and Gender," *Ethics* 105 (1994): 23–43, p. 31.

question whether a “reasonable comprehensive view” is or is not compatible with “justice as fairness.” One strong piece of evidence is that a “reasonable doctrine” is defined as “an exercise of theoretical reason: it covers the major religious, philosophical, and moral aspects of human life in a more or less consistent and coherent manner” (*PL*, p. 59). It is also an exercise in practical reason in “singling out which values to count as especially significant and how to balance them when they conflict. . . . Finally, a third feature is that while a reasonable comprehensive view is not necessarily fixed and unchanging, it normally belongs to, or draws upon, a tradition of thought and doctrine” (*PL*, p. 59). Rawls emphasizes that the definition is intended to be broadly inclusive. He says, rather mysteriously, that we are to “avoid excluding doctrines as unreasonable without strong grounds based on clear aspects of the reasonable itself” (*PL*, p. 59). I think this means that a doctrine is to be regarded as reasonable unless it can be proved to be unreasonable. And in this context it does not appear that its reaching conclusions at variance with those of justice as fairness is to be taken as proof of unreasonableness. Indeed, Rawls specifically denies that we are to take an independent view of the content of doctrines. “Political liberalism counts many familiar and traditional doctrines—religious, philosophical, and moral—as reasonable even though we could not seriously entertain them for ourselves” (*PL*, pp. 59–60).

These criteria are either formal or, we might say, sociological. To be reasonable, a comprehensive view must have the right scope and be cast in the right form. Beyond that, it looks as if any view that has stood the test of time will be given the nod. What Rawls is especially keen to avoid is apparently the exercise of independent judgment on the reasonableness of some view. If I were to say, for example, that I simply cannot see how any reasonable person could believe the tenets of the Mormon religion, that would be ruled out of court as inconsistent with the irenic spirit of “political liberalism.” There is, I suggest, more than whiff here of the idea found in (among others) Charles Larmore, according to which treating people with “equal respect” is supposed to entail treating their ideas with respect even if we privately find them extremely silly.³³

33. See Charles E. Larmore, *Patterns of Moral Complexity* (Cambridge: Cambridge University Press, 1987), p. 64. The notion that wants, however fantastic the beliefs on which they depend, should count as good currency in public policy underlies the argument of Rabbi Kimche for the granting of planning permission for erection of an *eruv* (“twenty foot posts connected near the top with fishing line”) to enable some orthodox Jews in northwest London to do some things on the Sabbath that they would otherwise (according to their beliefs) be unable to do. “You don’t understand, you won’t understand it, and, quite honestly, you don’t need to understand it. The point is we want it, we consider it important, and we ask you to respect that” (Calvin Trillin, “Drawing the Line,” *New Yorker* [December 12, 1994], pp. 50–62, p. 62).

In addition to Rawls's definition of a "reasonable comprehensive view," we can appeal to the whole structure of *Political Liberalism* to support the claim that he does not intend to foreclose by definitional fiat the question of the compatibility of reasonable views with "justice as fairness." As we shall see below (Sec. XII), Rawls lists a number of "reasonable comprehensive views" and then proceeds to make a case for the compatibility of each with "justice as fairness." This whole exercise would be otiose if it were conceptually impossible for a reasonable view to fail to be compatible with Rawlsian justice.

A final piece of evidence is that Rawls actually considers the possibility that compatibility might not obtain. Thus, he asks, "What if it turns out that the principles of justice as fairness cannot gain the support of reasonable doctrines, so that the case for stability fails?" (*PL*, p. 65). Rawls's answer is, "We should have to see whether acceptable changes in the principles of justice would achieve stability; or indeed whether stability could obtain for any democratic conception" (*PL*, p. 66). This response raises difficulties of its own.³⁴ But it at any rate makes it clear that there is not intended to be anything in the criteria of a "reasonable comprehensive view" that will guarantee the conformity of all such views with Rawls's principles of justice.

IX. SALVATION AND JUSTICE: TWO RESPONSES

The problem of stability with which Rawls is concerned is that of gaining free assent among reasonable people to the same principles of justice. But why should that be taken to acquire an "overlapping consensus" on them? Rawls rigs the argument by saying that the condition of stability is that "the reasonable doctrines endorse the political conception, each from its own point of view" (*PL*, p. 134). But it is people, not doctrines, that go around endorsing conceptions (or anything else). Rawls's way of putting it tacitly assumes the very point that is at issue: it presupposes that people can endorse principles of justice only if their "comprehensive view" endorses (in other words, entails or supports) them.

What makes *Political Liberalism* a puzzling text is that we find Rawls saying that reasonable people endorse the principles of justice (regardless of the content of their "comprehensive view") but then in the next breath claiming that it is their "comprehensive view" that

34. "Acceptable" would be redundant here if it meant "acceptable to the parties." It suggests an external criterion of acceptability, and this suggestion is reinforced by the stipulation that it must be a "democratic conception" around which an alternative consensus is to be built. Acceptable, then, to whom? The only answer seems to be "Rawls." But, once dislodged from his own theory, how could Rawls decide what to accept and how would he hope to persuade others to observe the same bounds of acceptability?

must endorse them. I can best illustrate this by taking up one key question: What can be said by a Rawlsian to somebody who claims that “outside the church is no salvation” and deduces from this the legitimacy of making conformity to the church in question compulsory? Rawls discussed this question in one of the earliest papers setting out his revised position, “The Idea of an Overlapping Consensus.”³⁵ He returned to it in the later article I have drawn on already, “The Domain of the Political and Overlapping Consensus,” and gave a different answer. In the second, he gave the response I advocate, “that the conclusion is unreasonable: it proposes to use the public’s political power . . . to enforce a view affecting constitutional essentials about which citizens as reasonable persons, given the burdens of reason, are bound to differ uncompromisingly in judgment.”³⁶ Rawls’s earlier reaction to somebody claiming that outside the church is no salvation was that “we may have no alternative but to deny this, and to assert the kind of thing we had hoped to avoid”—for example, to “assert in some form the doctrine of free religious faith that supports equal liberty of conscience.”³⁷ On this account, we have to say that somebody who holds a religious view that leads to illiberal implications is simply wrong rather than saying that, whether his view is right or wrong, it is unreasonable to force it on others.

In a forerunner of this article, which antedated *Political Liberalism* by several years, I noticed that Rawls had two different solutions to the same problem. Rather than simply pointing to an inconsistency in his views, however, I suggested that he had bounced back in “The Domain of the Political” from his earlier pessimism about the efficacy of the argument from the requirements of agreement on reasonable terms. This thought was encouraged by Rawls’s explicit repudiation in the later article of his earlier answer. Thus, he said in “The Domain of the Political” that “a reply from within an alternative comprehensive view—the kind of reply we should like to avoid in political discussion—would say that the doctrine in question is incorrect and rests on a misapprehension of the divine nature.” In contrast, the reply that he currently favored, he said, “does not say that [the] doctrine . . . is not true. Rather, it says that it is unreasonable to use the public’s political power to enforce it.”³⁸

Rawls continued by saying that we may wish to say a doctrine is incorrect as well as saying that it would be unreasonable to enforce it. (Presumably, affirming my own “comprehensive view” must commit

35. John Rawls, “The Idea of an Overlapping Consensus,” *Oxford Journal of Legal Studies* 7 (1987): 1–25.

36. Rawls, “The Domain of the Political,” p. 243.

37. Rawls, “The Idea of an Overlapping Consensus,” p. 14.

38. Rawls, “The Domain of the Political,” p. 243.

me to saying that other views inconsistent with it are false.) But he added the crucial point that saying a certain doctrine is true is equally compatible with saying it should not be enforced. "Indeed, it is vital to the idea of political liberalism that we may with perfect consistency hold that it would be unreasonable to use political power to enforce our own comprehensive religious, philosophical or moral views—views which we must, of course, affirm as true or reasonable (or at least as not unreasonable)."³⁹

Unfortunately, this attempt to save Rawls's consistency cannot survive the publication of *Political Liberalism*, where both the earlier and later versions reappear (the earlier on p. 152 and the later on p. 138). Even more curiously, in his further discussion of the earlier answer, Rawls interpolates a few new sentences that contradict it. The case he envisages is, he says, that of "rationalist believers who contend that these beliefs [about the route to salvation] are open to and can be fully established by reason" (*PL*, pp. 152–53). Now, such people apparently accept the criterion of reasonable agreement but simply maintain that everybody could reasonably accept their own views. Thus, "the believers simply deny what we have called 'the fact of reasonable pluralism.' So we say of the rationalist believers that they are mistaken in denying that fact; but we need not say that their religious beliefs are not true, since to deny that religious beliefs can be publicly and fully established by reason is not to say they are not true" (*PL*, p. 153).

This amounts to a statement of the later view. Yet Rawls immediately follows it by reverting to the earlier one, saying that "of course, we do not believe the doctrine believers here assert, and this is shown in what we do. Even if we do not, say, hold some form of the doctrine of free religious faith that supports equal liberty of conscience, our actions nevertheless imply that we believe the concern for salvation does not require anything incompatible with that liberty" (*PL*, p. 153). Yet what Rawls said just before this would suggest that, even if I myself held the view in question about the means to salvation, I should accept that the "fact of reasonable pluralism" makes it inadmissible as the charter of a whole society, since it can reasonably be denied by others.

Contrary to what Rawls contends, then, there is no need to reject the view that "outside the church is no salvation." What is to be rejected is the repressive conclusion drawn from it, and that is done by rejecting the claim that nobody could reasonably deny the truth of the doctrine. Rawls could respond by suggesting that every "comprehensive view" includes not only the content of the view but also some belief about

39. *Ibid.* There is something endearingly Rawlsian about the last sentence's dying fall.

the epistemological status of that content. Then, when he says that “we do not believe the doctrine believers here assert,” what he could be taken to mean is that we do not believe the epistemological aspect of the doctrine, though we might not wish to deny its content. Even if we took this line, though, it still would not follow that “our actions” in insisting on liberty of conscience “imply that we believe the concern for salvation does not require anything incompatible with that liberty.” For we would not be rejecting the doctrine that “outside the church is no salvation.” Rather, we would be making the familiar liberal point that ensuring salvation (according to one particular view of what that requires) is not an appropriate job for a state. The basis for this is the idea that any particular view about the means to salvation can reasonably be rejected. And this is, of course, an implication of the “fact of pluralism,” understood not as a brute fact about the actual existence of disagreement but as a claim about the irreducible variety of reasonable views under conditions in which freedom of thought is established.

X. IS DOGMATISM REASONABLE?

I have argued that the stability (as defined by Rawls) of “justice as fairness” does not require an “overlapping consensus” on it among “reasonable comprehensive views.” In this section and the next I want to see if there is any way of reinstating a connection between stability and “overlapping consensus.” Let me begin by returning to the idea that every “comprehensive view” might be taken to have two parts. One would be its content; the other would be a belief about its epistemic status. I shall dichotomize the range of possible epistemological views into those that can be squared with “the fact of pluralism” and those that cannot. People holding a view of the first kind accept that others may reasonably reject their first-order view. People holding a view of the second kind believe that no reasonable person could reject their first-order view—that only the invincibly ignorant or willfully blind could do so. I shall call “comprehensive views” incorporating the first view “skeptical” and those incorporating the second view “dogmatic.”

It will then be true that Rawlsian stability depends on an “overlapping consensus” among “reasonable comprehensive views” in the following sense: either a “comprehensive view” will have to include a skeptical second-order component or its first-order content will have to be compatible with Rawls’s principles of justice. Now, Rawls appears to maintain that it is unreasonable to deny “the fact of pluralism.” If so, it follows that all “reasonable comprehensive views” must be skeptical at the second-order level. The stability of Rawlsian justice is therefore guaranteed.

Although we are here employing the vocabulary of “overlapping consensus,” it is clear that all we are doing with it is recycling the first-

stage argument. Before, we said that reasonable people will accept the implications of “the fact of pluralism,” and thus refrain from pressing for their “comprehensive view” to be given a privileged position where “constitutional essentials” and the “basic structures of society” are concerned. Now we say instead that all “reasonable comprehensive views” have a recognition of “the fact of pluralism” built into them. The point is, obviously, the same both ways.

I do not put forward the two-level conception of a “comprehensive view” as an exegesis of Rawls’s own ideas. My point is that, even if we were to restate the first-stage argument in the terminology of “overlapping consensus,” this would not reinstate the necessity of Rawls’s own second-stage program. The discussion may, however, suggest a way in which one might defend Rawls’s own understanding of the conditions of stability. I have said that he appears to think it unreasonable to deny “the fact of pluralism.” But if he were to get cold feet about this assertion, regarding it as too controversial to form an element in the “political conception,” it would follow that dogmatists could count as reasonable. Then the only way of ensuring Rawlsian stability would be “overlapping consensus” at the first-order level.

It cannot be denied that Rawls exhibits a great deal of nervousness about making the “political conception” depend on an epistemological doctrine. He says, rather mysteriously, that “being reasonable is not an epistemological idea (though it has epistemological elements)” (*PL*, p. 62). The contrast is with “a political ideal of democratic citizenship that includes the idea of public reason” (*PL*, p. 62). Moreover, he denies that the ideas about “the burdens of judgment” that are needed by the “political conception” are skeptical. By this he says he means that they do not rest on some philosophical theory such as that of Descartes or Hume which concludes that “the necessary conditions of knowledge can never be satisfied” (*PL*, p. 63). Political liberalism, Rawls says, does not “question the possible truth of affirmations of faith.” But he goes on immediately to set out what I am calling skepticism: “the practical impossibility of reaching reasonable and workable political agreement in judgment on the truth of comprehensive doctrines” (*PL*, p. 63).

Rawls’s insistence that “the fact of pluralism” is not an epistemological doctrine and does not rest on any philosophical theory might lead us to doubt that it really does amount to what I am calling skepticism. Stephen Mulhall and Adam Swift suggest that “Rawls’s point about the burdens of reason is a sociological rather than a conceptual or logical one.” However, they do not deny that Rawls’s claims amount to “quasi-scepticism:” they simply argue that the evidence he offers does not entitle him to it. It is, they say, “really founded upon an empirical belief that the fact of pluralism is permanent rather than upon an irrefutable demonstration that it is intrinsically impossible to

establish reasonable agreement on comprehensive doctrines.”⁴⁰ But how many irrefutable demonstrations have ever been made in the entire history of western philosophy? Evidence of persistent failure to establish any “comprehensive view” by means appealing solely to human reason is surely quite a good basis for making the judgment that it would be unreasonable for anybody to claim an exception for any particular “comprehensive view.”

There is nothing original in drawing implications from the fact of persistent disagreement. Thus, John Stuart Mill, in his “Inaugural Address to the University of St. Andrews,” said that “diversity of opinion among men of equal ability and who have taken equal pains to arrive at the truth . . . should of itself be a warning to a conscientious teacher that he has no right to impose his opinion authoritatively upon a youthful mind.”⁴¹ Again, this is not an “irrefutable demonstration.” The premise that people who possess equal abilities and have made equal efforts have failed to reach agreement could be conceded, but the conclusion might still be drawn that the true doctrine ought to be instilled authoritatively from a tender age. But that does not prevent Mill’s premise from providing a sound reason for accepting his conclusion.

Rawls commits himself on occasion to what looks a pretty definite endorsement of the proposition that it is unreasonable to reject “the fact of pluralism,” understood as an epistemological claim rather than as a merely empirical one. Thus, in the course of making the case for the “political conception,” he says, “Since there are many reasonable doctrines, the idea of the reasonable does not require us, or others, to believe any specific reasonable doctrine, though we may do so” (*PL*, p. 60). This would seem to license the conclusion Rawls needs, that it is unreasonable to seek to make even a reasonable “comprehensive view” into the charter of a whole society. But let us suppose that his predominant view is that skepticism (as I define it) is too controversial to form part of the “political conception.” Then, the first-stage argument is no longer a self-contained basis for stability, and the burden has to shift to “overlapping consensus.”

XI. LIBERAL DEMOCRATIC STABILITY

I now wish to move on to a different way of arriving at an interest in an “overlapping consensus.” Let us suppose that dogmatism is inherently unreasonable. Nevertheless, we know all too well that dogmatists exist.

40. Stephen Mulhall and Adam Swift, *Liberals and Communitarians* (Oxford: Blackwell, 1992), p. 224. I am indebted to a conversation with Adam Swift for clarification of this point.

41. *Mill’s Essays on Literature and Society*, ed. J. B. Schneewind (New York: Collier Macmillan, 1965), p. 399.

If their “comprehensive views” are not compatible with Rawlsian principles, they will not freely accept institutions based on them. This is no challenge to Rawlsian stability, since that is defined as requiring agreement by the reasonable. If there are too many such dogmatists, they will threaten Hobbesian stability, but (as many examples show) Hobbesian stability is consistent with widespread disbelief in the moral foundations of the regime. What we need to be far more concerned about, however, is something lying between Rawlsian stability and Hobbesian stability. For easy reference I shall call it “liberal democratic stability.”

In Section III, I quoted Rawls as saying that the maintenance of a liberal democratic regime is incompatible with the need to repress more than a small fraction of the population. A good illustration of the truth of this is provided by the recent history of Northern Ireland, where a movement (the IRA) supported by only a little over one-tenth of the population (the average Sinn Fein vote) provided at the very least a widely accepted excuse for the suspension of key elements of liberal democracy. Since Rawlsian justice is an essentially liberal democratic notion, its maintenance is threatened by dogmatists whose “comprehensive view” does not endorse it. Notice that it makes no difference to this conclusion whether or not we say that dogmatism is inherently unreasonable, or whether the “comprehensive views” of dogmatists are reasonable or not. The point is simply that people who reject the “constitutional essentials” of liberal democracy in a dogmatic way represent a threat to the maintenance of liberal democratic institutions.

I said earlier that I accept Rawls’s claim that his new ideas stem from a concern for stability. I have argued, however, that Rawlsian stability does not need “overlapping consensus” unless “the fact of pluralism” is regarded as too controversial to found the “political conception.” We can now see another way in which Rawls could have arrived at a concern for the existence of “overlapping consensus.” This is its role in preventing dogmatists (even if we say they are unreasonable) from undermining liberal democracy and hence Rawlsian justice, which may be regarded as one specification of liberal democratic ideas.

In accepting Rawls’s own account of the origin of his new ideas, I also accepted his claim that (contrary to widespread assumptions) he was not driven to them by anything external to the theory of justice itself. However, if there is an external cause, I feel little doubt that it is to be found in the rise of religious dogmatism (or at any rate an increase in the salience of political demands derived from it) in the United States.⁴² We might perhaps say, though, that this is not com-

42. William Galston suggests that “as late as 1975 . . . the view of most Americans who regarded themselves as evangelical Christians” would have endorsed Jerry Falwell’s

pletely dissociated from the popular explanation of Rawls's new views as a response to "communitarian" criticisms, if we regard the "communitarian" movement as an academically attenuated form of the impulse to assert the political relevance of religion.

I have pointed out that, if we are worried about the threat to liberal-democratic stability posed by dogmatists who reject key elements of Rawlsian justice, it does not make any difference whether their first-order views are reasonable or not. Nevertheless, we might well think that there is little chance of "unreasonable and irrational, and even mad, comprehensive doctrines" supporting liberal democratic institutions. This could lead us to focus on the compatibility of "reasonable comprehensive views" with Rawlsian justice. Liberal democratic stability would still, of course, be potentially endangered by dogmatists with unreasonable first-order views. But Rawls could say about them (as he does) that "the problem is to contain them so that they do not undermine the unity and justice of society" (*PL*, pp. xvi–xvii).

XII. FIVE CANDIDATES FOR "OVERLAPPING CONSENSUS"

I have suggested two ways in which we might arrive at an interest in the compatibility of widely held "reasonable comprehensive doctrines" with "justice as fairness." I shall now ask how convincing are Rawls's attempts to show that his "political conception" forms a "module" (*PL*, p. 144) in a wide variety of "reasonable comprehensive doctrines." At one point, Rawls suggests grounds for optimism stemming from the way in which his principles are derived (via the original position) by abstracting from the content of different "comprehensive doctrines" (*PL*, p. 24, n. 27). But this does nothing to make the principles the locus of an "overlapping consensus." Thomists, say, might agree that Rawls's principles are what they would agree to if denied the knowledge of God, but quite properly deny that this somehow makes the principles an element in Thomism, which is based on theistic premises.

We need a list of what Rawls takes the "reasonable comprehensive views" to be. Before publishing *Political Liberalism*, he gave different lists in different places. It is an index of the rather weak editorial control that Rawls exercised in putting together the book that he does not take the opportunity to consolidate these into one list but presents two, each embedded in extracts from a different earlier paper. In one place he says that we may think of the "overlapping consensus" as

assertion that "preachers are not called to be politicians but soul winners." By 1980, however, he adds, "Falwell and his followers had reversed course and plunged headlong into the world" (William G. Galston, *Liberal Purposes: Goods, Virtues, and Diversity in the Liberal State* [New York: Cambridge University Press, 1991], p. 272). The sudden emergence of the doctrine of "overlapping consensus" in 1985 would fit neatly into this chronology.

containing “three views: one affirms the political conception because its religious doctrine and account of free faith lead to principles of toleration and underwrite the fundamental liberties of a constitutional regime; while the second view affirms the political conception on the basis of a comprehensive liberal moral doctrine such as those of Kant or Mill.” The third is described by Rawls as a “pluralist view.” It includes “the political values formulated by a freestanding political conception of justice” but also “a large family of nonpolitical values,” and all of these are to be “balanced against one another” (*PL*, p. 145). In the other account, “the religious doctrine with its account of free faith” is still the first view mentioned. Rawls now says, however, that the second item in the earlier list was “Kant’s moral philosophy with its ideal of autonomy,” whereas (as we have just seen) the second view was in fact the “comprehensive liberal doctrines” of Kant or Mill. In effect, the old second view is now bifurcated, but “for Mill’s view we substitute the utilitarianism of Bentham and Sidgwick” (*PL*, p. 169). The remaining view (the third in the earlier list and now the fourth) is the “pluralist account of the realms of values” (*PL*, p. 170).⁴³

Since Kant and Mill have almost nothing in common except that they are both in some sense liberal theorists, Rawls is right to separate them. He is also right to distinguish Mill from what he calls the “classical doctrine” of utilitarianism (*PL*, p. 170). Putting the two accounts together, then, there would seem to be five candidates for “overlapping consensus”: a religious doctrine of “free faith,” Mill, Kant, utilitarianism, and a mixed bag of values including justice.

Of the five candidates, there can be no doubt that the Kantian is the most plausible. Indeed, it is the only one for which Rawls makes the claim that it provides a “deductive basis of the political conception” (*PL*, p. 169). (The most he claims for the others is that they can in some way be construed so as to accommodate the principles of justice.) This may throw light on a question it is natural to ask about Rawls’s list. If, as he now says, the doctrine of *A Theory of Justice* constituted a “comprehensive view,” why does it not appear as one of the candidates for membership in an “overlapping consensus”? The answer is, I suggest, that it does—but in the guise of “[Kant’s] view or . . . a view sufficiently similar to it” (*PL*, p. 169).⁴⁴ This, it may be noted, supports

43. The first account follows “Justice as Fairness,” p. 250; the second account roughly follows the text of an unpublished public lecture, “The Idea of Free Public Reason” (Inaugural Abraham Melden Lectures, Department of Philosophy, University of California at Irvine, February 20 and March 1, 1990, typescript, p. 6).

44. The chapter “Political Constructivism” in *Political Liberalism* contains an extended discussion of the relation between Kant and the kind of constructivism that enters into the first-stage argument leading to the “political conception.” On Rawls’s account, Kant’s theory is in a number of ways too deep, and therefore too controversial,

my surmise in Section VI that it is its reliance on the “Kantian interpretation” for the stability of justice that has led Rawls to say that *A Theory of Justice* comprises a “comprehensive view.”

Let us accept that there is an interpretation of Kant, or perhaps we should say a revision of Kant that retains enough to count as Kantian, from which Rawls’s principles of justice can be deduced. We must also observe that there are other interpretations or developments of Kant’s ideas with implications very far from those of Rawls’s principles. It could be (and has been) maintained, for example, that the injunction not to treat people as means only has the implication of ruling out any kind of redistribution. The normal response of a philosopher faced with this kind of challenge is to confront it directly, and aim to refute it. This, however, would involve precisely the kind of substantive argument that Rawls says must be eschewed by anyone engaged in an ecumenical project such as his own.

Let us next take up “the utilitarianism of Bentham and Sidgwick” and the “comprehensive liberal doctrine” of Mill. On the basis of considerations such as our limited knowledge and the limits on the complexity of rules, “the utilitarian [may] think a political conception of justice liberal in content a satisfactory, perhaps the best, workable approximation to what the principle of utility, all things tallied up, would require” (*PL*, p. 170). Mill’s relation to Rawlsian “justice as fairness” is not specified, but I take it that Rawls would want to say again that Mill’s practical conclusions approximate his own, though they are arrived at on the basis of a conception of the good as autonomy.

I cannot undertake here any serious attempt to ask how far Rawls’s claims are justified. Let me simply say that at the very least the question is far more open than one would gather from Rawls. With the right “stylized facts,” utilitarianism can be made to generate something like Rawls’s first principle; with different ones it can lead to the rejection of both liberty and democracy.⁴⁵ The same goes for the second principle of justice: those who are impressed by the importance of stable expectations for prosperity, like Bentham, will argue (like Hayek today) for a low “safety net” but no more; other utilitarians will derive something not too far from the “difference principle” from the diminishing marginal utility of money.⁴⁶ As far as Mill is concerned, a con-

to form the basis of a whole society’s adopting “justice as fairness” (*PL*, pp. 99–101). It is, nevertheless, pretty clear that he endorses this theory (on his own interpretation of it) himself.

45. Sir James Fitzjames Stephen notoriously contradicted *On Liberty* at all points on the basis of the claim that Mill was a bad utilitarian. See his *Liberty, Equality, Fraternity* (London: Smith, Elder, 1873).

46. Thus, A. C. Pigou, in *The Economics of Welfare* 4th ed. (London: Macmillan, 1932), wrote that “the direct good of transference and the indirect evil resulting from

cern for autonomy does plausibly lead to support for Rawls's first principle. But it seems plain that a concern with autonomy is compatible with a wide variety of views about distribution. One quite natural conclusion is the libertarian one that maximizing autonomy requires that the state enforces contracts but does not redistribute the resources that people acquire under a regime of free exchange. There are more egalitarian alternatives, but we clearly cannot say that those who put a high priority on autonomy must endorse Rawls's principles of justice.

The next view that Rawls includes in his list of candidates is, he says, "not systematically unified: besides the political values formulated by a freestanding political conception of justice, it includes a large family of nonpolitical values" (*PL*, p. 145). Later, Rawls adds that "each domain of value has . . . its own free-standing account" (*PL*, p. 170). People are to decide what to do "by balancing judgments that support the great values of the political against whatever values normally conflict with them in a well-ordered democratic regime" (*PL*, p. 170). Rawls asserts that in this balancing process "the political conception is affirmed" when it conflicts with other values (*PL*, p. 170). This is, of course, correct as an assertion of what has to be the case for this "mixed conception" to form part of the "overlapping consensus." But is there any good reason for believing it to be true?

An easy way of making it true would be to pick up the idea that "each domain of values has its own free-standing account" and point out that the free-standing account of justice (i.e., the first-stage argument) actually has as a theorem the moral priority of justice over competing values. If the validity of this free-standing account is accepted, then, there is no room for any question about the priority of the demands of justice when these come into conflict with those arising from other values. This is, however, no more than to reaffirm the self-sufficiency of the "political conception." If that is accepted, the whole exercise within which the "mixed conception" plays a part is redundant.

Since the context is one in which stability is taken to depend on the existence of "overlapping consensus," we must assume that the self-sufficiency of the "political conception" has somehow been ruled out. Then the problem that Rawls is describing sounds very much like the problem of "congruence" in *A Theory of Justice*, minus the Kantian ace in the hole. The problem then became, we may recall, one of a balance of motives (*TJ*, p. 57: see Sec. V above). And Rawls was then

a diminished dividend are both finite quantities; and the correct formal answer to our question is that economic welfare is best promoted by a minimum standard raised to such a level that the direct good resulting from the transference of the marginal pound transferred to the poor just balances the indirect evil brought about by the consequent reduction of the dividend" (p. 761).

quite depressive about the likelihood of what he now calls the “great values of the political” (civil peace and social harmony) reliably coming out on top when in conflict with other values. It is hard to see why anything that has happened since 1971 should make him more optimistic—if anything, the opposite.

We are now left with the “comprehensive view” that, according to Rawls, “affirms the political conception because its religious doctrine and account of free faith lead to a principle of toleration and underwrite the fundamental liberties of a constitutional regime” (*PL*, p. 145). The “account of free faith” is glossed in a footnote (with reference to Locke), the key notion in it being that “only faith and inward sincerity gain our salvation and acceptance with God” (*PL*, p. 145, n. 12).

The stipulation that only religious views that incorporate “an account of free faith” are to count as reasonable is hard to square with Rawls’s hospitable definition of a “reasonable comprehensive view.” For, as we saw in Section VIII, Rawls insists that we are not to make our own judgments of reasonableness but, rather, are to accept as reasonable “many familiar and traditional doctrines—religious, philosophical, and moral” that we “could not seriously entertain . . . ourselves” (*PL*, pp. 59–60). If we are prepared to accept (as Rawls clearly intends) that the three major monotheistic religions are to count as reasonable, I do not see how we can say that it would be unreasonable for their adherents to read the books on which they are based as rejecting the “account of free faith.” The Gods of Moses, Mahomet, and Saint Paul (Jesus is perhaps an exception) do not sound like liberals. Rawls closes the gap by an act of sheer bravado, saying “here I shall suppose—perhaps too optimistically—that, except for certain kinds of fundamentalism, all the main historical religions admit of [an account of free faith] and thus may be seen as reasonable comprehensive doctrines” (*PL*, p. 170).

Even if we concede that an account of free faith can be extracted from all the main historical religions, Rawls’s stipulation entails that they will count as reasonable only when interpreted in accordance with this account. Thus, we are faced with two unappealing alternatives. One is to follow Rawls’s content-free definition of reasonableness, in which case it will turn out that many adherents of mainstream religions have reasonable views that reject “free faith.” The other is to follow Rawls in saying that only those versions that incorporate “free faith” are to count as reasonable, thus introducing exactly the kind of content-based criterion of reasonableness that he said was to be avoided.

Let us assume, either way, that Rawls is prepared to make claims to membership in the “overlapping consensus” only for those religious views that incorporate “free faith.” The case for compatibility with the principles of justice still seems to me rather weak. No doubt “an ac-

count of free faith" can be extended until it underwrites Rawls's first principle of justice, and this process has been significant in the development of civil liberties. But there is nothing to guarantee this. "Free faith" assumes that God is a liberal only to this extent: that he does not value professions of religious belief that are coerced. But it does not rule out the suppression of heterodox religious faiths on the ground that they may tempt the faithful from the true path.⁴⁷ Nor does it rule out the forcible conversion of one generation in the hope (which history would largely justify) that following generations will voluntarily accept the dominant religion. Even worse, "free faith" does nothing to defend free action: God may still value the prohibition of what the dominant religion defines as sinful acts. Moreover, religions naturally tend to support theocracy rather than democracy, since the implementation of the religion's demands is most likely to be faithfully carried out by the religion's own hierarchy. Thus, there is no reason in general for expecting religions, even if they accept the "doctrine of free faith," to endorse the equal civil and political rights that make up Rawls's first principle of justice. As far as the second principle of justice is concerned, it is significant that Rawls does not even attempt to show that religious views incorporating the "account of free faith" must underwrite it.

XIII. "OVERLAPPING CONSENSUS": AN ASSESSMENT

We are interested in overlapping consensus, I have suggested, for two reasons. First, it connects with liberal democratic stability; second, it connects with Rawlsian stability. The connections are different in the two cases, however. For liberal democratic stability we need only agreement (to an adequate degree) on something like Rawls's first principle of justice. Perhaps at some point lack of agreement on the legitimacy of socioeconomic inequality threatens the stability of liberal democracy. But experience shows that a liberal democracy in which there is widespread consensus on the "rules of the game" can contain a remarkable amount of conflict over the kinds of question addressed by Rawls's second principle. For Rawlsian stability, we need agreement on both Rawls's first and second principles: the "political conception" is said by Rawls to have the same content as in *A Theory of Justice*, so the agreement required to make justice stable must likewise extend to both principles.

If the scope of consensus called for by liberal democratic stability is smaller than that demanded by Rawlsian stability, the former is more

47. *A fortiori*, it does not rule out the suppression of "books that would be sufficient, if read, to shake the faith of an otherwise orthodox population." Jeremy Waldron, *Liberal Rights: Collected Papers, 1981–1991* (New York: Cambridge University Press, 1993), p. 109.

demanding in the range of people among whom the consensus has to exist. Rawlsian stability can withstand rejection by the unreasonable, whereas liberal democratic stability is liable to be undermined by the sheer existence of a significant minority that rejects the fundamental elements of liberal democracy, whether reasonably or not.

Let us assume that all skeptics are to count as reasonable and that skepticism underwrites Rawls's principles of justice. Then skeptics are no problem for either Rawlsian stability or liberal democratic stability. For liberal democratic stability, we need agreement among dogmatists on (roughly) Rawls's first principle of justice. The first four of Rawls's five "reasonable comprehensive views" are quite promising in this respect. The last is, I have suggested, a good deal more problematic. The greatest threat to liberal democracy probably comes, however, from those who combine dogmatism with the rejection of "an account of free faith." Let us call them fanatics. It is a robust generalization that the existence of a large bloc of fanatics makes liberal democracy impossible. Whether the fanatics take control (as in Iran) or are thwarted by force (as in Algeria), liberal democracy is impossible.

Rawlsian stability does not have to concern itself with the last category of dogmatists, if we take Rawls as saying that any religious doctrine not including "an account of free faith" is *ipso facto* unreasonable. We can then say, as before, that Rawls's first principle of justice is at least tolerably well established among the first four "reasonable comprehensive views," though its standing with the fifth is shaky. It is when we come to the second principle of justice that the strategy of "overlapping consensus" collapses. The first three "reasonable comprehensive views" can be made to yield an approximation of the "difference principle," but at least as easily lend themselves to other conclusions about the legitimate grounds of socioeconomic inequality. The implications of the fourth, "mixed," conception presumably depend on the "values" put into it and their relative weights: it is hard to see how more could be said about something so vacuous.

This leaves religious "comprehensive views." No doubt Christianity can be given an egalitarian slant. (We may think of leveling Protestant sects and the "preferential option for the poor" of Liberation Theology.) But mainstream Christian denominations have always tolerated socioeconomic inequalities (including, in the past, slavery) vastly in excess of anything that could be justified by Rawls's "difference principle." Islam and Judaism embrace a similar spread of views, while Hindu and Confucian systems are inegalitarian to the core in a way that no monotheistic religion can be. It is therefore almost inconceivable that Rawls's second principle of justice could be presented as an inescapable implication of all the major religions.

Let me enlarge on the last point. Rawls, and others sympathetic to his project, tend to suggest that important progress has been made

if it can be shown that there is one version of each “comprehensive view” that can be squared (more or less) with the principles of justice.⁴⁸ But this would be significant only if there were some good reason for attributing to such an interpretation a competitive edge simply in virtue of its compatibility with Rawls’s principles. We might then anticipate that over time this interpretation would gain ground and eventually approach hegemonic status. (I assume that we are here looking for some ground independent of the merits of the interpretation, as seen from within the “comprehensive views.” Of course, if it could be shown for each “comprehensive view” that the Rawlsian interpretation is the most rationally compelling one, the problem would be solved triumphantly. But Rawls does not ever suggest that he believes this, nor are the moves he makes appropriate to showing it.)

I can see no reason of the required kind. Consider first the case of somebody who does not accept the free-standing argument for the principles of justice outlined above in Section VII. It is hard to see why such a person should be attracted to an interpretation of his “comprehensive view” simply in virtue of its consilience with Rawls’s principles of justice. Now take the case of somebody who is persuaded to adopt Rawlsian principles on the basis of the “political conception.” This already tells him that there are limits to the extent to which he can press for his “comprehensive view” to be translated into public policy, at any rate where “constitutional essentials” and the “basic structure of society” are at issue. It does not tell him that he has any reason for modifying his “comprehensive view” so that it too endorses Rawls’s principles; all it says is that, given the “fact of pluralism,” he must accept that others could reasonably reject it, and this entails that it would be unfair to make it the charter of the whole society.

Perhaps we should be looking not for reasons but for causes.⁴⁹ We might surmise that, although there is no logical inconsistency between holding a “comprehensive view” with anti-Rawlsian implications and simultaneously endorsing Rawlsian principles on the basis of the first-

48. See S. A. Lloyd, “Relativizing Rawls,” *Chicago-Kent Law Review* 69 (1994): 709–35, esp. 733–34.

49. The only sustained attempt I know of to suggest a mechanism tending to push “comprehensive views” in a Rawlsian direction is that made by Joshua Cohen in his “A More Democratic Liberalism,” *Michigan Law Review* 92 (1994): 1503–46, on 1530–36. He suggests that living in a society that instantiates Rawlsian principles creates a “pressure on views” (p. 1532) to come into line. As Cohen concedes, however (on p. 1553), this process—a causal one, in my terms—is actually neutral in itself. It implies that in a society with Nazi institutions, the “learning” (p. 1535) process will shift “comprehensive views” in a Nazi direction, and that a society in which net incomes become more disparate will see an increase in antiegalitarian views. Germany in the past and the United States today provide some support for the Cohen hypothesis; but the examples dramatize my point that there is no inherently pro-Rawlsian tendency in the mechanism.

stage argument for them, it is nevertheless uncomfortable. The tension could be relieved, without abandoning the “comprehensive view,” by shifting to an interpretation of it compatible with Rawlsian principles. (It could also, of course, be relieved by rejecting the first-stage argument, but let us be optimistic.)

Here, at last, we have a mechanism that confers an advantage, independent of intrinsic merit, on an interpretation of a “comprehensive view” that fits Rawlsian principles of justice, thus giving significance to the possibility of producing one interpretation (even if it is only one of many) that is consistent with these principles. Unfortunately, however, it recruits to the “overlapping consensus” only those for whom membership is irrelevant to stability. For the mechanism can operate only on those who already accept the principles of justice in virtue of their derivation within the “political conception.” And these people, as Rawls says, already have Scanlonian motivation for respecting the demands of justice.

What are the implications? In an initial statement of the themes of *Political Liberalism*, Rawls says that “the two principles [of justice] express an egalitarian form of liberalism. . . . All these elements are still in place, as they were in *Theory*; and so is the basis of the argument for them” (*PL*, pp. 6–7). Rawls adds to this assertion of continuity a rather querulous footnote which runs as follows: “I make this comment since some have thought that my working out the ideas of political liberalism meant giving up the egalitarian conception of *Theory*. I am not aware of any revisions that imply such a change and think the surmise has no basis” (*PL*, p. 7, n. 6). The obvious basis for the surmise is that, although the principles are unchanged at the first stage, the second principle is abandoned at the second stage. And, since the principles of justice can be reasonable for a society only if they can be endorsed from within each reasonable comprehensive view, the implication is surely that the second principle has to be sacrificed.

It may, of course, be doubted if any alternative socioeconomic distributive principle could form the subject of an “overlapping consensus.” But Rawls could respond to this by simply dropping the second principle and saying that the “constitutional essentials” on which agreement is required do not extend beyond the first principle. It is significant that, in a subsequent publication, Rawls appears to accept that just liberal societies need not incorporate the egalitarian commitment he emphasized in *Political Liberalism*.⁵⁰ This is, it seems to me, no more than a logical working out of the conditions of stability as Rawls now conceives them. If the second principle cannot be sustained at the second stage, it must be excised from the first stage too.

50. John Rawls, “The Law of Peoples,” in *On Human Rights*, ed. Stephen Shute and Susan Hurley (New York: Basic, 1993), pp. 41–82, p. 51.

This pessimistic assessment of the prospects of Rawlsian stability is not necessary, of course, if we are prepared to say that dogmatism is inherently unreasonable. Then, as long as Rawls's principles of justice are derivable from skepticism, we can conclude that reasonable people must accept them. As I have tried to show, Rawls is remarkably ambivalent about the reasonableness of dogmatism. On one side we have the most recent strand of *Political Liberalism*, which is most prominent in the chapter which bears the title "The Idea of Public Reason" (*PL*, pp. 212–54). Rawls says all the right things here about the need for norms of public discourse that require citizens to respect the "burdens of reason." The obvious implication (which, as we have seen, Rawls occasionally acknowledges) is that they cannot reasonably press claims based on their own "comprehensive views" that others are bound to reject. Yet Rawls does not draw the conclusion that "overlapping consensus" is unnecessary for stability. The other position is still there, and is indeed a good deal more central to the book. According to this, we need an "overlapping consensus" not among reasonable people but among "reasonable comprehensive views." But this makes sense as a condition of consensus on the principles of justice only if it is possible for somebody to be a reasonable dogmatist.

XIV. CONCLUSION

I have not in this essay made any special effort to conceal my own views on the questions discussed, but I have aimed to keep the focus on Rawls's own work. Let me conclude briefly by expressing my own conviction that the idea of "overlapping consensus" is thoroughly misconceived. In the relations between the public and the private there are, it seems to me, two extremes to be avoided. One—exemplified by fascism, extreme nationalism, Jacobinism, and bolshevism—dissolves the private into the public, abolishing the claims of private judgment altogether. The other extreme is represented by philosophical anarchism, and makes private judgment sovereign.⁵¹ Rawls's idea that the principles of justice must be endorsed from within each "reasonable comprehensive view" on its own terms is in essence a capitulation to this style of thinking.⁵² There is a middle ground, which alone

51. See Robert Paul Wolff, *In Defense of Anarchism* (New York: Harper & Row, 1970); and A. John Simmons, *On the Edge of Anarchy* (Princeton, N.J.: Princeton University Press, 1993).

52. What exactly is each "reasonable comprehensive view" to endorse? Samuel Scheffler takes Rawls to mean that "overlapping consensus" requires endorsement of the whole "political conception of justice"—what I have called here (in Sec. VII) the first-stage argument for the principles of justice. See Samuel Scheffler, "The Appeal of Political Liberalism," *Ethics* 105 (1994): 4–22. He rightly describes this as an "extraordinarily strong" demand, and points out that it cannot possibly be met by utilitarianism. For the first-stage argument is essentially that of *A Theory of Justice*, which is an avowedly

makes liberal democratic politics secure, and which is absolutely required to underwrite large public projects, including the kind of economic reconstruction called for by Rawls's second principle. This combines skepticism on conceptions of the good or "comprehensive views" with strong conclusions about the requirements of justice. Rawls's first-stage argument provides a model of a theory occupying this middle ground. Without endorsing every detail, I do want to insist that it is the right kind of thing.

Montesquieu said that a bad book by a famous author causes a lot of trouble.⁵³ In Rawls's case, the trouble arises because it is almost universally assumed that, if he now says there were fundamental flaws in *A Theory of Justice*, he must at any rate be right about that. Since there is also a widespread feeling that *Political Liberalism* does not succeed in fulfilling its stated task, the conclusion is naturally drawn that the whole Rawlsian project is fatally flawed. I have argued that Rawls's sweeping recantation is uncalled-for, and that the failure of *Political Liberalism* does not discredit *A Theory of Justice*.⁵⁴ I believe that, as time goes on, *A Theory of Justice* will stand out with increasing clarity as by far the most significant contribution to political philosophy produced in this century. Only one thing threatens to obscure that achievement: the publication of *Political Liberalism*.

antiutilitarian work. The same might, I believe, be said of the other candidates, with the sole exception of the "Kantian" one. But, as I noted in Sec. XII, this is the only one for which Rawls is prepared to advance the claim that "the political conception . . . can . . . be derived" from it (*PL*, p. 169). Scheffler also correctly points out that the derivation of the "political conception" from within each "comprehensive view" is unnecessary for Rawls's purposes, because stability needs consensus on the principles of justice but not on a single set of grounds for accepting them (p. 14). Surely, these are reasons for not attributing the view in question to Rawls.

53. "Je supplie le lecteur de me pardonner l'ennui mortel que tant de citations doivent lui donner: je serais plus court, si je ne trouvais toujours devant moi le livre . . . de M. l'abbé Dubos. Rien ne recule plus le progrès des connaissances qu'un mauvais ouvrage d'un auteur célèbre, parce qu'avant d'instruire il faut commencer par détromper" (Charles Secondat, Baron Montesquieu, *De l'Esprit des lois*, ed. Gonzague Truc [Paris: Editions Garnier Frères, 1956], vol. 2, bk. 3, chap. 15, p. 317).

54. I do, of course, think that Rawls is right to be dissatisfied with chap. 9 of *A Theory of Justice*; but here his dissatisfaction does not seem to me to go anywhere deep enough. The only thing to do with it is to follow the course followed virtually unanimously by commentators on *A Theory of Justice* and forget about it.