Political Constructivism in Rawl's Political Liberalism

George Klosko

The American Political Science Review, Vol. 91, No. 3 (Sep., 1997), 635-646.

Political Constructivism in Rawls's *Political Liberalism*
GEORGE KLOSKO   *University of Virginia*

*I*n Political Liberalism, *John Rawls employs a distinctive method of "political constructivism" to establish his well-known principles of justice, arguing that his principles are suited to bridge the ineradicable pluralism of liberal societies and so to ground an "overlapping consensus." Setting aside the question of whether Rawls's method supports his principles, I argue that he does not adequately defend reliance on this particular method rather than alternatives. If the goal of Rawls's "political" philosophy is to derive principles that are able to overcome liberal pluralism, then another and simpler method should be employed. The "method of convergence" would develop liberal principles directly from the convergence of comprehensive views in existing societies, and so give rise to quite different moral principles.*

In *Political Liberalism*, John Rawls (1993) employs a distinctive method of "political constructivism" to establish his well-known principles of justice. He argues that his principles are suited to bridge the ineradicable pluralism of liberal societies and so to ground an "overlapping consensus." I will not question his claim that his method supports his principles, but I do not believe Rawls adequately defends his reliance on this particular method rather than alternatives. If the goal of Rawls's "political" philosophy is to derive principles that are able to overcome liberal pluralism, then another and simpler method should be employed, which would also give rise to quite different moral principles.

For reasons of space, I concentrate on the first part of this two-part argument. I begin by reviewing Rawls's conception of political liberalism and his method of political constructivism. Next, I contrast his method with an alternative, to which I refer as the method of convergence, and suggest the alternative moral principles this other method would produce. I then argue for the superiority of this alternative method on the basis of detailed criticisms of Rawls's method. Finally, I examine two possible criticisms of my argument and present a brief conclusion.

RAWLS'S POLITICAL CONSTRUCTIVISM

In *Political Liberalism* Rawls argues that fundamental disagreements in liberal societies are unbridgeable. Because people who deliberate about moral and political issues emphasize different aspects of questions and employ different methods of investigation, they will come to different and irreconcilable answers (Rawls 1993, 54–8). Rawls calls this fact of human reason the "burdens of judgment," and he argues that a diversity of doctrines is "a permanent feature" of liberal public

culture, not an accident of history, destined to pass away (Rawls 1993, 216–7).

Political philosophy contributes to the stability of liberal societies by providing a basis for agreement where agreement is possible. Toward this end, its subject matter is restricted in two important ways. First, the range of issues with which it deals is circumscribed. Rather than address all questions that might concern citizens, political liberalism is confined to what Rawls calls society's "basic structure," its "main political, social, and economic institutions, and how these fit together into one unified system of social cooperation" (Rawls 1993, 11). Rawls includes in the basic structure certain ideas and values; it encompasses not only the "framework of basic institutions" but also "the principles, standards, and precepts that apply to it, as well as how these norms are to be expressed in the character and attitudes of the members of society who realize its ideals" (Rawls 1993, 11–2). In contrast to a "political" view that deals with basic structures, Rawls describes a moral view as "general" if it addresses a wide range of subjects.

Second, Rawls describes a moral view as "comprehensive" if it contains answers to a full range of difficult and controversial questions, for example, what is of value in human life and ideals of personal character. A view is "fully comprehensive" if it encompasses "all recognized values and virtues within one rather precisely articulated system" (Rawls 1993, 13). In contrast to comprehensive views, a political conception is articulated without reference to wider philosophical background. In order to be compatible with the range of comprehensive views found in contemporary liberal societies, a political conception should be as far as possible independent of the opposing and conflicting philosophical and religious doctrines that citizens affirm (Rawls 1993, 10–1).

In order to attain the support of adherents of different comprehensive views, political liberalism attempts to generate an "overlapping consensus." Briefly, Rawls argues that, in spite of their important differences, the comprehensive views in liberal societies are able to agree in regard to a range of issues and principles bearing on the smooth functioning of society's basic structures. Though people will support this range of principles for different reasons, with those of each person stemming from his or her own compre-

hensive view, the area of agreement throughout society will be sufficiently broad to contribute to stability.

In addition to the necessary content, an overlapping consensus is accepted by its adherents in a particular way. Rawls contrasts the overlapping consensus with a modus vivendi, which is conceived on the model of a truce, the outcome of political bargaining. Because people adhere to a modus vivendi for self-interested reasons, its terms reflect the balance of power between contending factions and so are subject to renegotiation as the balance shifts (Rawls 1993, 147). In contrast, an overlapping consensus is a moral conception to which people subscribe for moral reasons. Its adherents will not withdraw their support if the relative strength of their view increases and eventually becomes dominant (Rawls 1993, 147–8). Because Rawls believes that the balance of power on which a modus vivendi rests can be precarious, he views the distinction between modus vivendi and overlapping consensus as important to social stability.

In spite of obvious respects in which *Political Liberalism* differs from Rawls's earlier *Theory of Justice*, he believes that much of the content of the latter work is retained in the former (Rawls 1993, xiv–xvi). Especially important for our purposes, he argues that "justice as fairness," the celebrated principles of justice defended in *Theory of Justice*, constitutes the content of the overlapping consensus that should exist in a liberal society.[1] The two principles are familiar and need not be discussed in detail here (Rawls 1971, 60–1, 302–3).[2] The first guarantees equal liberty for all members of society. The second has two parts. It guarantees, first, that positions in society are to be open to everyone and distributed according to fair equality of opportunity and, second, that economic inequalities must benefit the least advantaged members of society, Rawls's famous "difference principle."

The method of "political constructivism" through which the principles are derived is complex. As in *Theory of Justice*, the principles of justice are the outcome of a process of choice, conducted in the original position, behind a veil of ignorance. The original position is a "device of representation" to help focus our moral ideas: "It models what we regard— here and now—as fair conditions under which the representatives of free and equal citizens are to specify the terms of social cooperation in the case of the structure of society" (Rawls 1993, 25–6). Given problems caused by the pluralism of liberal societies, the principles arrived at must be acceptable to adherents of society's different comprehensive views. Accordingly, argument must be from "intuitive ideas" that Rawls believes are deeply rooted in liberal culture and to which adherents of different comprehensive views must

subscribe. Rawls describes public culture as comprised of "the political institutions of a constitutional regime and the public traditions of their interpretation (including those of the judiciary), as well as historic texts and doctrines that are common knowledge" (Rawls 1993, 13–14).

The two specific intuitive ideas on which Rawls focuses are a view of society as a fair system of cooperation and a conception of the person as possessing two moral powers, concerning abilities to form and revise one's own conception of the good and to live on fair terms of cooperation with others. Though Rawls does not describe in detail how the intuitive ideas tie in with the choice of principles in the original position, it is clear that the intuitive ideas are represented by central features of the original position and the deliberations of the representative individuals. Because these particular conceptions of the person and of society are built into the structure of the original position, selection of principles of justice under these conditions identifies them as the most suitable principles for free and equal citizens, possessing the two moral powers, who seek to live on fair terms of cooperation with others.

Because the principles of justice must be acceptable to adherents of different comprehensive views, Rawls argues that justice as fairness is a "freestanding view." This means that it is not constructed on the basis of the existing comprehensive views found in society but, rather, through the intuitive ideas. Rawls divides the process of construction into two stages. In the first, the principles are constructed without reference to existing comprehensive views, although, apparently, inhabitants of the original position can be aware of the existence of particular views in particular societies, but not the ones to which they themselves subscribe. Rawls says of the principles of justice: "Their content is not affected in any way by the particular comprehensive doctrines that may exist in society" (Rawls 1993, 141). Once the principles of justice are chosen in the first stage, they are reviewed in regard to whether they would be acceptable to proponents of society's comprehensive view, or as Rawls terms this, in regard to their contributions to stability.[3] If the principles are lacking in this respect, suitable adjustments may be necessary (Rawls 1993, 65–6).

For the sake of argument, we can assume that Rawls's principles of justice would be the outcome of the method he describes.[4] But this goes only part of the way toward making his case. Basic features of the selection process itself must be justified. In *Theory of Justice*, Rawls notes that different versions of the conditions under which principles of justice are chosen will yield different sets of principles: "We may conjecture that for each traditional conception of justice there exists an interpretation of the initial situation in

---

[1] Rawls is somewhat ambiguous as to whether his aim is the best possible principles of justice or simply acceptable principles, under the assumption that more than one set is acceptable; see Rawls 1993, 164; 1987, 6–7; 1980, 565, 569. In either case, however, it is safe to say that justice as fairness could constitute the core of an overlapping consensus and that Rawls is not aware of alternative principles that work better.

[2] For Rawls's latest statement of the principles, see 1993, 5–6.

[3] For Rawls's use of "stability," see 1989, 239 n12. For the concept of stability elsewhere in Rawls's works and problems with it, see Klosko 1994. Stability is discussed below in the conclusion.

[4] An excellent account of a possible argument for this outcome is given in Cohen 1993.

which its principles are the preferred solution" (Rawls 1971, 121). In order to avoid vicious circularity, Rawls must show that the features he presents rather than others should characterize the choice situation. The fact that he never makes this argument in detail is a serious shortcoming of *Political Liberalism*—though this does not necessarily mean that reasons for these features do not exist. I will attempt to piece together Rawls's defense of the method, or the defense he would give if pressed, from points he makes in different contexts in *Political Liberalism* and in some of his articles leading up to it.[5]

The two main problems with Rawls's method on which I will focus are (1) his concentration on two— and only two—basic, intuitive ideas in liberal public culture and (2) his account of justice as fairness as a freestanding view. I will argue that these features of Rawls's method are not defensible in reference to the specific tasks of political liberalism. Rather, as I have noted, a simpler, alternative method is better suited to these tasks. The method of convergence is without these objectionable features of Rawls's method.

CONSTRUCTIVISM AND CONVERGENCE

In order to consider Rawls's method, we must inquire into exactly what political liberalism is to accomplish. Assessment of means depends heavily on the end. There is, however, a problem here, in that Rawls sets two goals for political liberalism, which do not necessarily coincide. The "first fundamental question" he presents about political justice in a democratic society is as follows: What is the most appropriate conception of justice for specifying the fair terms of social cooperation between citizens regarded as free and equal, and as fully cooperating members of society over a complete life ... (Rawls 1993, 3)?

The aim here is to identify the *best* possible principles, with the criteria clearly normative. The second question concerns what Rawls calls "toleration understood in a general way" (Rawls 1993, 3). Rawls wishes to find a means of bridging the multiplicity of reasonable doctrines that arise in liberal society. His aim here is practical, to identify principles on which adherents of different doctrines can agree, so as to contribute to social stability (Rawls 1993, 3–4). I will not discuss the additional requirement that the principles of justice must be "liberal" (Rawls 1993, 6; 1987, 17; 1989, 241). We should note that, given the diversity of the liberal tradition, this requirement should still leave a great deal of latitude. Another important issue I will not address concerns the nature of the desired consensus in liberal societies. Different sorts of agreements can be envisioned, especially encompassing different percentages of the population (different degrees of popular consensus), with narrow agreement perhaps being more substantial in terms of both percentage of belief-

systems agreed on (extent of issue agreement) and intensity of commitment to shared beliefs.[6]

Rawls's conception of how his principles are justified is in keeping with the practical task:

> Justification is addressed to others who disagree with us, and therefore it must always proceed from some consensus, that is, from premises that we and others publicly recognize as true. . . .
>
> Thus the aim of justice as fairness, as a political conception is practical, and not metaphysical or epistemological. That is, it presents itself not as a conception of justice that is true, but one that can serve as a basis of informed and willing political agreement between citizens viewed as free and equal persons (Rawls 1985, 229–30; similarly 1987, 6; 1989, 250)

He writes similarly in *Political Liberalism*: "The aim of justice as fairness, then, is practical: it presents itself as a conception of justice that may be shared by citizens as a basis of a reasoned, informed, and willing political agreement. It expresses their shared and public political reason" (Rawls 1993, 9).

Though his remarks on justification emphasize the practical side of his project, Rawls places greater weight on the normative. He addresses this in the first of the two fundamental questions and in the first of the two stages of political construction. Yet, the need to respond satisfactorily to the practical question places limits on possible answers to the normative. Rawls recognizes this in his requirement that construction be from intuitive ideas in the political culture and the need for a second stage in the construction process. But I believe he does not go far enough. The need to address the practical question should make Rawls's political liberalism more political than he himself acknowledges.

An immediate difficulty confronts an attempt to find the most appropriate conception of justice for a pluralistic society. Because liberal societies are torn by deep disagreements over fundamental values, Rawls recognizes the likelihood of disagreements over principles of justice. He resorts to the intuitive ideas to find a way around this problem. The intuitive ideas are intended to be generally subscribed to in ways that principles of justice are not. Accordingly, Rawls describes them as "public and shared ideas" (Rawls 1993, 90). However, because of the burdens of judgment, it seems unlikely that adherents of conflicting comprehensive views will readily agree that a specific conception of justice is best suited to free and equal people

---

[5] The main articles on which I draw are Rawls 1980, 1985, 1987, and 1989. Also important are Rawls 1982 (reprinted as Lecture 8 of *Political Liberalism*) and 1988.

[6] Rawls apparently prefers a somewhat narrower popular consensus, as is evident in his insistence that the overlapping consensus be between "reasonable" views (1993, 36, 63–4). I will not explore different ways in which the variables in regard to consensus could play out, though we should note that Rawls defines "reasonable" quite narrowly and so may well exclude many inhabitants of liberal societies from the necessary consensus; for criteria of "reasonable" comprehensive views, see 1993, 58–62; and below, in the next section; for related criticisms along these lines, see Wenar 1995. Given the tasks of "political" philosophy, I believe it is preferable to construe "reasonable" more broadly. The only people who are not "reasonable" are those whose comprehensive views would prevent them from coexisting peacefully with other members of society. See Rawls 1971, 216–21, on tolerating the intolerant.

who have to live together. Among the many issues over which adherents of different views will probably disagree are the precise characteristics of free and equal persons. "Freedom" and "equality" are "essentially contested" concepts (Connolly 1983, chapter 1; Gallie 1955–56). In the absence of strong evidence to the contrary, there is little reason to believe liberal citizens will agree more readily about these issues than about other aspects of their moral views. Indeed, one reason Rawls advances for his method of construction is the existence of conflicting strands in liberal societies concerning how the values of freedom and equality should be reflected in the rights of citizens (Rawls 1993, 4–5).

Rawls addresses the possibility that principles derived in the first stage of construction will not be generally accepted in the second. He notes that this would require "acceptable changes" in the principles to bring them into accord with society's comprehensive views (Rawls 1993, 66). But he sets this problem aside, assuming—"on the basis of a number of plausible considerations"—that it will not arise (Rawls 1993, 66). If we are less sanguine about this problem, we can see that Rawls glosses over serious difficulties, especially the need to take appropriate steps in the first stage of construction to make sure the resultant principles will fit with society's views. Because of his emphasis on the pluralism of liberal societies and the burdens of judgment, the burden of proof is clearly on Rawls to show that justice as fairness would fit. He notes that his claims can be "verified only by actually elaborating a political conception of justice and exhibiting the way in which it will be supported" (Rawls 1987, 6–7; similarly, 1993, 15). But Rawls never attempts to meet this challenge. One commentator characterizes his view as "wildly optimistic."[7] *Political Liberalism*—and the articles leading up to it—are strikingly lacking in detailed examination of the comprehensive views of actual liberal societies.[8]

I do not contend that a fit between Rawls's principles and liberal culture is impossible. But Rawls's particular method of construction, pursuing as it does the best possible principles that can be drawn from selected intuitive ideas, is defensible only if his principles will fit. Because of the strong possibility they will not, a more advisable procedure would focus on the need to generate principles that will fit and then select from these the ones which are normatively preferable. In other words, if the goal is to find serviceable principles, then the order of the two stages of construction should be reversed.

Because of the need to find principles that will be acceptable to different comprehensive views, a preferred method would aim first at generating a set of principles that are as uncontroversial as possible. This

in turn creates a strong presumption against relying on controversial notions, including a specific conception of the free and equal person, in their construction. The alternative method I propose begins with an attempt to find areas of agreement between existing comprehensive views. Once such an area is identified, we would proceed to derive principles with the most robust possible normative content. Throughout this paper, I will concentrate on the first of these tasks, leaving aside complex issues concerning argument in the second. I assume that a variety of normative approaches could be employed, some of which are drawn from Rawls. The first stage of this alternative method, to which I refer indifferently as convergence or the method of convergence, can be represented visually as a Venn diagram, with a highlighted area of agreement—or overlap—between different comprehensive views, in spite of their other differences.

The convergence method attempts to avoid controversy as far as possible. Notion X is by definition uncontroversial, or as uncontroversial as possible, if it is adhered to by the different groups in liberal society, even if members of each group interpret it differently from their own perspectives. Methodological rules concerning the need to avoid controversy are presented in Rawls's articles. Thus, in "Idea of an Overlapping Consensus," he writes: "The question is: what is the least that must be asserted; and if it must be asserted, what is its least controversial form?" (Rawls 1987, 8). Along similar lines: "The aspects of our view that we assert should not go beyond what is necessary for the political aim of consensus" (Rawls 1987, 14; similarly 1985, 229–30). Though we find nothing as strong as the first of these quotations in *Political Liberalism*, Rawls restates the second: "We do not put forward more of our comprehensive view than we think needed or useful for the political aim of consensus" (Rawls 1993, 153).[9] Because of the need to avoid controversy, Rawls argues in "Justice as Fairness: Political Not Metaphysical" that one must "draw *solely* upon basic intuitive ideas" from the public culture (Rawls 1985, 225; emphasis added). In *Political Liberalism*, language concerning the need to rely solely on such ideas is not found, though Rawls does not explain whether this represents a change in his views or whether, in the process of construction, one can employ material beyond shared intuitive ideas—and where the line in this regard should be drawn.[10]

The success of the method of convergence depends on the existence of widespread agreement in liberal societies. But evidence for such agreement can be provided, while it is apparent that different groups are more likely to agree on the relatively bland principles such evidence supports than on the distinctively Kantian principles Rawls advances.

The existence of an important area of agreement is

---

[7] Jones 1995, 526–7. To explain Rawls's belief that the views he regards as "reasonable" doctrines will be accepted as the content of the overlapping consensus, Jones argues that Rawls runs together two senses of "reasonable." Rawls bases his view of the overlapping consensus "on the wrong sort of reasonableness," that of the doctrine to be accepted rather than that of the people who are to accept it.
[8] For Rawls's empirical investigations of liberal societies in the articles leading up to *Political Liberalism*, see Klosko 1993.

[9] Aside from the insertion of a paragraph, Rawls 1993, 152, follows 1987, 14. Rawls does not indicate whether he still accepts the stronger injunction on 1987, 8.
[10] I do not find language of "solely" in Rawls 1987, 6–7; but it is used in 1989, 240.

argued by Kurt Baier (1989) in a well-known article. Although he believes that consensus on a conception of justice of the kind Rawls posits does not exist in contemporary American society, Baier contends that a different one does, a "constitutional consensus." He describes this as general agreement on "procedures for making and interpreting law and, where that agreement is insufficiently deep to end disagreement, on the selection of persons whose adjudication is accepted as authoritative." As Baier notes, the constitutional consensus performs important functions akin to those Rawls assigns to the overlapping consensus, promoting stability in society and fostering essential civic virtues of "tolerance, respect and reciprocity." He also claims that agreement on institutions and procedures is supported by agreement on a general principle of justice associated with the proper functioning of institutions.[11]

Rawls's account (1993, 149 n15, 158 n24) of constitutional consensus is similar to that of Baier, and Rawls appears to accept Baier's contention that one presently exists in the United States (Rawls 1993, 149). Like Baier, Rawls believes that in American society there is greater agreement on political and civil rights than on matters of economic distribution. He includes the former but not the latter in his view of constitutional consensus, which he describes as consensus on "only certain fundamental procedural political principles for the constitution" (Rawls 1993, 149).

> While there is agreement on certain basic political rights and liberties—on the right to vote and freedom of political speech and association, and whatever else is required for the electoral and legislative procedures of democracy— there is disagreement among those holding liberal principles as to the more exact content and boundaries of these rights and liberties, as well as on what further rights and liberties are to be counted as basic and so merit legal if not constitutional protection (Rawls 1993, 159).

Even though Rawls notes these important limitations, it bears emphasis that the moral content of constitutional consensus is significant, emphasizing respect for certain rights, while willingness to put one's own preferences aside and accede to the results of decision procedures is an important political virtue and one central to the liberal tradition. Rawls also notes that the practice of constitutional politics fosters "the cooperative virtues of political life," reasonableness, a sense of fairness, and willingness to compromise (Rawls 1993, 163).

Despite these virtues, and the respects in which constitutional consensus is similar to overlapping consensus, Rawls views the former as inadequate. His reasons are as we have seen. Constitutional consensus falls short in not addressing questions of economic distribution, that is, the subject of Rawls's second principle of justice. This is its most obvious deficiency. Constitutional consensus also has a weaker conception of rights, the content of which is unacceptably subject to "the shifting circumstances" of political bargaining (Rawls 1993, 161). Rawls also objects to the means

through which constitutional consensus is derived. Its precepts do not stem from shared ideas of the person and the nature of society but are "accepted simply as principles" and so lack moral depth (Rawls 1993, 158). For these reasons, then, Rawls argues for overlapping consensus and his method of constructivism.[12]

Once again, however, Rawls's reasoning can be criticized. Though overlapping consensus is doubtless desirable—and constitutional consensus not morally ideal—Rawls's movement to this higher form of consensus (Rawls 1993, 164–8) is difficult to defend, absent convincing evidence for its existence or at least its plausibility. But as we have seen, Rawls does not provide this. His criticisms of the way constitutional consensus is derived presuppose the superiority of his method. We will examine Rawls's defense of his method in the next section.

To close out this section, I present brief empirical evidence that a constitutional consensus, unlike an overlapping consensus, currently exists in American society. This evidence is obviously only a sketch and requires further elaboration. But it should suffice to establish at least the plausibility that a constitutional consensus exists in American society at the present time. As Baier notes, the core of constitutional consensus is agreement on procedures for making laws and means of adjudication. The evidence that interests me bears on a certain type of agreement, on what David Easton calls "diffuse political support." Easton describes this as "a reservoir of favorable attitudes or good will that helps members to accept or tolerate outputs to which they are opposed or the effects of which they see as damaging to their wants" (Easton 1965, 273) In other words, diffuse support correlates closely with support for decision-making mechanisms—"procedures and arrangements through which demands are negotiated and compromised" (Easton 1965, 272)—without immediate regard to specific decisions.

A series of national surveys indicate strong support for the American political system, for the decision-making mechanisms in the United States. This was confirmed by a telephone survey of 1,443 people conducted in 1992 under the direction of John Hibbing and Elizabeth Theiss-Morse (1995). They inquired into support for different political institutions, opinions which they attempted to keep separate from those concerning people serving in the offices in question. The results are striking. Congress was approved or strongly approved by 88% of respondents, the Supreme Court by 94%, and the Presidency by 96% (Hibbing and Theiss-Morse 1995, 44–5). Only 8% of respondents disapproved of the "basic constitutional structure of the U.S. government." Only nine individuals (.06%) strongly disapproved (p. 59).[13] James Hunter and Carl Bowman (1996), in their 2,047 face-to-face interviews

---

[11] Baier 1989; all references are to 775. For Rawls's response to Baier, see 1993, 158–68.

[12] Rawls's arguments against constitutional consensus are discussed in more detail in the conclusion.
[13] These results obtained despite the declining levels of "political trust" found in American National Election Studies. A main contention of Hibbing and Theiss-Morse (1995, esp. chapter 5) is that people distinguish between political figures—the Washington sys-

conducted in early 1996 by the Gallup Organization, did not concentrate on having their respondents distinguish between support for the political system and for individuals officeholders, but their results are similar. In response to "Do you feel you should support our system of government?" The responses were 80% positive, 11% neither positive nor negative, and 9% negative; only 2% were strongly negative (Hunter and Bowman 1996, Table 12D).[14]

This evidence obviously only broaches the surface. But it suggests the likelihood of general acceptance of basic American political institutions and the decisions they produce. The method of convergence would focus on this, identifying it as an area of constitutional consensus, despite wide differences among citizens' comprehensive views. Once the relevant areas of agreement are identified in the first stage of this modified process of construction, one can argue in the second for different versions of specific principles it supports.[15] We will see in the next section, through detailed examination of Rawls's political constructivism, that the convergence method is preferable.

CRITIQUE OF POLITICAL CONSTRUCTIVISM

Fuller consideration of Rawls's political constructivism can center on the two problem areas noted in the first section. We will examine his ground for focusing on two, and only two, intuitive ideas and his claim that justice as fairness is a freestanding view.

Rawls is clearly aware of the method of convergence. At one point, he describes two ways in which principles can be constructed, using as his example an index of primary goods. The first way is "to look at the various comprehensive doctrines actually found in society and specify an index of such goods so as to be near to those doctrines' center of gravity, so to speak" (Rawls 1993, 39). This obviously corresponds to convergence. In contrast, justice as fairness "elaborates a political conception as a freestanding view . . . working from the fundamental idea of society as a fair system of cooperation and its companion ideas" (Rawls 1993, 40). Rawls's hope is not that his principles "are fair to comprehensive conceptions of the good associated with such doctrines, by striking a fair balance among them, but rather fair to free and equal citizens as those persons who have those conceptions" (Rawls 1993, 40; cf. 1985, 228; 1993, 8). Our questions, of course,

tem—and institutional structures, strongly disapproving of the former while strongly approving of the latter.

[14] The important study by Muller, Jukam, and Seligson (1982) presents similar findings and also provides data concerning diffuse support in West Germany and in Guadalajara, Mexico, and San Jose, Costa Rica.

[15] Discussion of specific principles the method would establish must be set aside for another paper. Very briefly, since the method identifies consensus on support for democratic political institutions, preferred principles will center on support for democracy and on strong preconditions for effective democracy, though in normative content, these must fall well short of Rawls's overlapping consensus; see below, in the concluding section.

concern why Rawls proceeds in this way and whether his procedure is justified.

Intuitive Ideas

Reliance on intuitive ideas, even ones that people hold, is not without costs. If a central aim of political liberalism is to address each group according to ideas to which it subscribes, then it is difficult to understand why Rawls believes in selecting from people's overall moral views two—and only two—intuitive ideas from which to generate principles. People should be more likely to accept ideas drawn directly from views they hold, as opposed to constructed implications, some attenuated implications, drawn from only some of their intuitive ideas. If these are truly *intuitive* ideas, people may not be aware of holding them and are unlikely to recognize that they are committed to their implications.[16]

Rawls would respond by appealing to the pluralism of liberal societies. The fact that people do not agree on fundamental matters of justice makes it necessary to organize public culture around particular focal points: "If we are to succeed in finding a basis for public agreement, we must find a way of organizing familiar ideas and principles into a conception of political justice that expresses those ideas and principles in a somewhat different way than before" (Rawls 1993, 9). "Since no political agreement on those disputed questions [concerning matters of justice] can reasonably be expected, we turn instead to the fundamental ideas we seem to share through the public political culture" (Rawls 1993, 150). Once again, justice as fairness focuses on free and equal persons cooperating in society.

But Rawls never presents evidence to support the claim that people agree on the intuitive ideas though disagreeing about substantive principles of justice. In "Kantian Constructivism in Moral Theory," he describes the ideas as "conjectured to be implicit" in liberal culture (Rawls 1980, 569). One must wonder if the ideas on which Rawls focuses are appropriate bases for popular consensus. He notes that the intuitive ideas chosen must be the most central possible; justice as fairness must be based on "more central fundamental ideas" than other conceptions (Rawls 1993, 167–8). But he does not demonstrate that this is true of his two basic ideas. He does not explain why public culture should be organized around them rather than others.

The lack of adequate explanation is especially troubling because Rawls's view of the person is not particularly plausible. Choice of intuitive ideas is limited by the circumstances of public culture. In order to fulfill their function, the particular ideas on which Rawls focuses must actually be basic features of public culture, common to its different comprehensive views.[17]

[16] On intuitive ideas, see Klosko 1993, 350.

[17] Since the practical goal is to derive principles that will fit with society's comprehensive views, it is possible that one can do this by employing intuitive ideas less central to public culture, as long as principles that will be accepted can be generated through them.

Central to the conception of the person Rawls employs is a strong measure of moral autonomy. He argues that the ability to examine one's moral beliefs and revise one's conception of the good "is not a means to but is an essential part of a determinate conception of the good."[18] He continues:

> The distinctive place in justice as fairness of this conception is that it enables us to view our final aims and loyalties in a way that realizes to the full extent one of the moral powers in terms of which persons are characterized in this political conception of justice (Rawls 1993, 314).

We can refer to the ability to revise one's conception of the good as "moral revisability."

Once again, it is striking that Rawls never explores liberal public culture in detail to demonstrate the centrality of this conception of the person. It is especially important that he do so, because religious conservatives do not place at the center of their view of the person the ability to revise and change one's conception of the good. For instance, according to James Hunter (1990), central to the moral outlook of religious conservatives is belief in "an external, definable, and transcendent authority." This has profound implications for their view of personal identity, which centers upon the need to pursue and live according to moral truth. Belief in such a moral standard "defines, at least in the abstract, a consistent, unchangeable measure of value, purpose, goodness, and identity, both personal and collective." "Orthodox communities order themselves, live by, and build upon the substance of shared commitment to transcendent truths and the moral traditions that uphold them" (Hunter 1990, 44, 126). According to Hunter, about 36% of Americans do not view the ability to think and reason as "the most important feature of human life," but the fact that man is created in God's image (Hunter 1994, 94). One person interviewed by Hunter said:

> God tells me what's right and what's wrong. I may attempt in a limited capacity to try to understand that, but I have to start off from the point that I am surrendering my personal intellect to God. If something doesn't make sense to me, that has no bearing on the reality of it or my obligation to respond to it (Hunter 1990, 121).[19]

Rawls is clear that the conception of the person he discusses is intended as a political conception, as opposed to a philosophical or metaphysical account. One sense in which ability to change one's view of the good is essential to a political conception of the person is that changes in what one values, even sudden and dramatic changes of the kind experienced by Saul of Tarsus on the road to Damascus, do not affect one's "public or institutional identity" or standing in society (Rawls 1993, 30–1). For instance, the fact that a person becomes a devout member of some religious sect rather than an atheist does not affect his rights as a citizen or her ability to hold property. We can concede that, in this sense, the ability to revise one's conception of the good probably is central to the conception of the person found in liberal political culture.

Rawls goes well beyond moral revisability in this sense, however, making it central to the choice of moral principles.[20] Because the two moral ideas are modeled by the constructivist procedure pursued in *Political Liberalism*, religious conservatives could well have strong doubts about Rawls's resultant principles of justice. It is open to Rawls to dismiss this criticism. Religious conservatives may not be "reasonable" in his sense, as that requires one to recognize the burdens of judgment (Rawls 1993, 54–66). Rawls argues that he is justified in excluding unreasonable doctrines from consideration because his theory is intended to produce an overlapping consensus of reasonable doctrines (Rawls 1993, 36, 63–4). For the sake of argument, we can grant that the views of religious conservatives are not reasonable in Rawls's sense. There are, however, obvious disadvantages to excluding such groups. A central purpose of the overlapping consensus is to promote social stability. At one point Rawls notes that he is optimistic in assuming that, aside from certain forms of fundamentalism, all the main historical religions are reasonable comprehensive views (Rawls 1993, 170). The problem, however, is the likelihood that views Rawls would classify as fundamentalism, that is, different forms of religious conservatism, are adhered to by about 20% of the U.S. population.[21] Moreover, as Leif Wenar points out, Rawls's conception of the person would be rejected by adherents of many nonreligious comprehensive views, such as followers of Bentham, Hume, and Hobbes (Wenar 1995, 50). Further reflection probably would identify additional groups.[22] In the

---

Though we cannot rule out this possibility, it is clearly an unlikely procedure, unless one could show how it would work. Rawls would have no easy time defending such a procedure as the best way to derive principles that are likely to be widely accepted.

[18] Though Rawls relies on a strong conception of autonomy, his view should be distinguished from the "comprehensive liberal theories" of Kant and Mill, which are explicitly based on controversial, strong notions of autonomy (Rawls 1993, 78, 199–200). Rawls's position is that his conception of autonomy is not rooted in a comprehensive view. It is found, rather, in intuitive ideas widely shared throughout society, though it appears that this contention would be difficult to sustain.

[19] Similar points about the clash between Rawls's view of the person and the religious views of many Americans are made by Galston (1989, 714; 1992, 130–1), Wenar (1995), and Exdell (1994, 447; I owe this last reference to Exdell).

[20] There is little doubt that Rawls goes beyond autonomy in the first sense (political autonomy) to identify autonomy in the stronger sense as central to liberal culture, but it is not clear that this plays an important role in his derivation of the principles of justice. See Rawls 1993, 311 n23; compare sec. 82 of *Theory of Justice*, which does not rely on autonomy in the strong sense. Thus, it is unclear why Rawls presents so strong and controversial a view of the person, despite the injunction to avoid controversy in construction, discussed above. For related criticisms of Rawls's view, see Wenar 1995.

[21] This is Hunter's estimate (1990, 159). Hunter and Bowman (1996) identify between 20% and 25% of the American population as Evangelical, which they define as someone (among other things) to whom at least two of the following apply: s/he has a personal relationship with God, believes scripture should be taken literally, and prays daily (Hunter and Bowman 1996, 52, 98 n13). Using a slightly different description, Jelen and Wilcox (1995, 37) identify 33% of the U.S. population as Evangelical. Exactly what constitutes an Evangelical Protestant need not be discussed here. It is unlikely that people falling under either description subscribe to Rawls's view of the person.

[22] Members of many other religious groups would probably also

absence of strong evidence to the contrary, then, it seems likely that Rawls's view would not be recognized by a large percentage of the American populace. Obviously, if social stability requires consensus, as many of these people as possible should be accommodated. This can be done immediately by drawing on ideas that they already accept, instead of principles derived from controversial intuitive ideas.

A Freestanding View

Questions concerning Rawls's use of intuitive ideas lead directly into the second main question about his method, his two-step procedure. Because justice as fairness is a freestanding view, its construction is in two stages. In the first, as we have seen, principles of justice are worked up from intuitive ideas latent in the public culture. Only in the second stage is the question of stability raised. What interests us here is that only at this stage do society's different comprehensive views come into play. As they are derived in the first stage, the principles' content "is not affected in any way by the particular comprehensive doctrines that may exist in society" (Rawls 1993, 141).

Rawls does not explain exactly why comprehensive views must be excluded in the first stage but not the second. For the most part, discussion of his method centers on contrasting it with other possible methods, though not defending its superiority.[23] It seems that Rawls's main argument is that convergence would make political liberalism "political in the wrong way." In two contexts in *Political Liberalism*, Rawls uses this language (1993, 39–40, 142). Though he does not elaborate on this problem in this work, Rawls presents more of his reasoning in "The Domain of the Political and Overlapping Consensus" (1989). In this article, which introduced the two-stage model, Rawls expresses "misgivings" that the view of overlapping consensus advanced in his previous articles had suggested that political philosophy is "political in the wrong way."[24] He developed the two-stage method to avoid this problem (see Rawls 1989, 250–1). Rawls writes:

> We must, however, be careful that a political conception is not political in the wrong way. It should aim to formulate a coherent view of the very great (moral) values applying to the political relationship and to set out a public basis of justification for free institutions in a manner accessible to free public reason. It must not be political [i] in the sense of merely specifying a workable compromise between

known and existing interests, nor political [ii] in looking to the particular comprehensive doctrines known to exist in society and in then being tailored to gain their allegiance (Rawls 1989, 250).

The reasons for rejecting the first sense are evident. As Rawls says in *Theory of Justice*, "to each according to his threat advantage is not a principle of justice" (1971, 141). But considerations against the second sense, which corresponds to the method of convergence, are less apparent. Because Rawls does not present his arguments clearly, I must attempt to reconstruct them.

Rawls apparently views a consensus based on convergence as inherently precarious. As his contrast between overlapping consensus and modus vivendi makes clear, he is worried about principles derived from agreements based on an existing balance of power, which are subject to renegotiation as the balance shifts. Hence it is likely that Rawls fears that if principles were chosen through negotiations between existing comprehensive views, then their substance "would be affected by the existing balance of political power between comprehensive doctrines" (Rawls 1993, 142). He obviously wishes to avoid such a bargaining situation, presumably because of concern about the stability of the resulting principles, which recalls the modus vivendi: "social consensus founded on self- or group interests, or on the outcome of political bargaining: social unity is only apparent, as its stability is contingent on circumstances remaining such as not to upset the fortunate convergence of interests" (p. 147).

If this is Rawls's main ground for preferring the two-step procedure, then it is unsatisfactory for three main reasons. First, the image of different people, or proponents of different groups, engaged in political bargaining recalls Rawls's initial presentation of his social contract theorizing, in "Justice as Fairness" (1958). This model was found to be unsatisfactory because people who are aware of their interests have no reason not to hold out for concessions.[25] One reason Rawls introduced the veil of ignorance in *Theory of Justice* was to get around this problem, as people who do not know their particular interests are not able to bargain in the usual sense (Rawls 1971, 139–41; also 1993, 23). Similar considerations should hold for the "bargaining" that would generate liberal principles in a pluralist society under the method of convergence. Throughout *Political Liberalism*, Rawls employs the veil of ignorance (especially 1993, 23–5), the existence of which should not be affected by whether the material used to construct liberal principles is drawn from selected intuitive ideas or society's different comprehensive views. A modified veil of ignorance can allow knowledge of the range of comprehensive views present in a given society while still excluding which one a given person holds. In deriving liberal principles from the area of overlap between comprehensive views through the method of convergence, representative individuals should work impartially and not be influenced by the particular comprehensive views to which they happen to subscribe. They

---

object to Rawls's conception of the person. It should be noted that these groups, and the ones mentioned in the text, are clearly *excluded* from Rawls's consensus, while, once again, he presents no evidence as to the percentage of the population that would be included. In contrast, the surveys cited above show far fewer members of society clearly excluded.

[23] See Rawls 1993, chapter 3, especially the list of topics to be discussed, on p. 89.

[24] Throughout *Political Liberalism*, Rawls draws relatively little on "Domain" in comparison, for example, to "The Idea of an Overlapping Consensus" (1987). But there is nothing in *Political Liberalism* to suggest that he has rejected its arguments. Barry (1995, 891–2) also uses "Domain" to flesh out certain points in *Political Liberalism*. The "previous articles" mentioned are, apparently, Rawls 1985 and 1987.

[25] For discussion of Rawls 1958, see Wolff 1977, chapter 5.

should not attempt to promote their own good but should work toward a compromise that would be fair to adherents of all comprehensive views. In regard to something akin to convergence, Rawls speaks of striking "a fair balance" among different comprehensive views (1993, 40). Accordingly, because some semblance of the veil of ignorance is retained with this method, proponents of different views should accept resultant principles as fair and not withdraw allegiance once they step out from behind the veil. In fearing that principles based on knowledge of the comprehensive views in society would be the result of undesirable bargaining, Rawls seems not to take into account the effects of the veil of ignorance, which, once again, the method of convergence does not discard.

Second, Rawls's exposition raises questions about the extent to which the two stages are in fact separate and so whether the content of the principles is actually "not affected in any way by the particular comprehensive doctrines that may exist in society" (Rawls 1993, 141). As it stands, this claim is incorrect. Derivation of the principles in the first stage is not ex nihilo. Intuitive ideas rooted in the public culture, of course, play an essential role in their construction. Rawls does not explain why, if construction of the principles is inherently tainted by exposure to comprehensive views, intuitive ideas drawn from existing comprehensive views are admissible but beyond this point not allowed. Excluding other, possibly relevant aspects of comprehensive views is also counterintuitive because of the practical aims of political constructivism. Principles derived in the first stage are not intended solely to be as just as possible; they must also fit with society's comprehensive views. As we have seen, at times Rawls notes this (1993, 65, 78). He also notes that if the principles selected in the first stage would not fit properly, "acceptable changes" must be made (1993, 65–6). Accordingly, the stages are not as separate as Rawls asserts, and so there are clear reasons not to exclude knowledge of everything but the two basic intuitive ideas from the first stage.

A third consideration concerns another possible reason for excluding comprehensive views. Central to Rawls's idea of "political" liberalism is that people should be able to endorse the principles from their own points of view, according to their own comprehensive views. Rawls writes: "To attain such a shared reason, the conception of justice should be, as far as possible, independent of the opposing and conflicting philosophical and religious doctrines that citizens affirm" (1993, 9). Part of what he means by public justification is that, in a liberal society, coercive public power should not be used in ways that citizens cannot affirm (1993, 68). Yet, this entails that justification of the use of power not be rooted in a particular comprehensive view. It does not require that construction of the relevant principles be fully independent of conflicting comprehensive views, as long as principles constructed by working from different existing comprehensive views can be justified to each citizen in his or her own terms, on the basis of his or her own comprehensive view.

It seems, then, that Rawls's reasons for fearing that

principles worked up from existing comprehensive views would be "political in the wrong way" can be countered. As long as the method of convergence preserves the veil of ignorance, so people are unaware of their own comprehensive views, it should be able to generate principles that can be generally endorsed—without the problems we have seen in working from selected intuitive ideas.

FURTHER ELABORATION AND CONCLUSION

Having discussed Rawls's method of political constructivism, contrasted it with the simpler method of convergence, and seen the alternative principles to which the latter gives rise, I will now consider two possible objections. One concerns the large degree of overlap between constitutional consensus and overlapping consensus; the other centers on implications of their areas of difference. To begin with the overlap, the principles that convergence develops, centering on support for the political system, are quite close to support for the "constitutional essentials" that Rawls's method derives as the central component of his conception of political justice (see Rawls 1993, 227–30). But if this is true, then it appears to blunt the foregoing criticisms of Rawls's method. Why criticize his method if it yields what we regard as essentially correct results?[26]

To this line of criticism, there is a response. The main focus of this paper is Rawls's method of political constructivism and how he justifies it. The fact that central components of the principles it develops are consistent with those that would stem from convergence does not provide support for the features that distinguish Rawls's method from convergence, for which, as we have seen, he does not provide strong grounds. We have seen that convergence allows a less cumbersome and less controversial way of deriving principles that center on support for the political system. Even though Rawls's method yields similar results, convergence would still appear to be preferable, for these reasons.

More important is the area in which Rawls's constructivism and convergence yield different principles. Though both methods develop principles centering on support for the political system, Rawls's method is required to derive the full content of his conception of justice, which also includes principles bearing on distribution, in Rawls's case, the difference principle.

I will not contest the claim that it is desirable to add economic principles to society's principles of justice (though, in a pluralistic society, not everyone will agree). Instead, we must ask if Rawls's method is the best way to develop these. He notes that the necessary consensus in society must be "consistent with plausibly realistic assumptions" (Rawls 1993, 149). We have seen

---

[26] I assume that what Rawls identifies as constitutional essentials (1993, 227–30) basically form the core of constitutional consensus, though the particulars of the essentials will differ between constitutional consensus and overlapping consensus (see 1993, 164–8, 227–30).

reasons to doubt certain assumptions underlying his employment of his method. Accordingly, there is little reason to think that principles derived in this particular way are those most suited to become the focus of widespread agreement throughout society. Once again, Rawls provides little reason to think that they are, and he does not demonstrate that they would come to be agreed upon as components of an overlapping consensus.

Empirical data provide little support for the difference principle. As David Miller (1992, 578) writes in a recent survey of empirical studies of attitudes about distributive justice, to the extent that people were seen to hold common views on the subject, they were not supportive of the difference principle. Instead, "the overwhelmingly popular choice was to maximize the average income subject to a floor constraint."[27] Miller notes strong support for desert "as a major criterion of income distribution" (1992, 580), though researchers found some disagreement among members of different economic classes on the extent to which questions of distribution should turn on talents and abilities (Miller 1992, 584–5). Accordingly, direct evidence of public culture does not support Rawls's preferred principles of distribution, while, once again, throughout this paper we have seen reasons to be skeptical about his method's ability to develop the most appropriate principles.[28]

The second criticism takes a different tack, centering on differences between constitutional consensus and overlapping consensus. If the method of convergence gives rise to only the former, and this can be shown to be inadequate to achieve central goals of liberal theory that the latter is able to accomplish, then Rawls's method perhaps can be defended as required by the goals of liberal theory. In order to be persuasive, an argument along these lines must provide a convincing account of how the areas of difference between overlapping consensus and constitutional consensus are not only desirable but also necessary. It must show that constitutional consensus falls short in crucial ways, so that remaining at that level is not a viable option. Though Rawls's ideas on this subject can be found in different contexts in *Political Liberalism*, they are somewhat unclear; once again, I will attempt to reconstruct them. His direct discussion of the contrast between the two kinds of consensus is largely unhelpful, because it is concerned to show that overlapping consensus is not "utopian" (1993, 168, 158), rather than that it is necessary.[29]

At first sight, it would be difficult for Rawls to show that constitutional consensus is obviously unsatisfactory. He notes that overlapping consensus is not necessary "for certain kinds of social unity and stability" and that constitutional consensus is "satisfactory for less demanding purposes and easier to obtain" than overlapping consensus (Rawls 1993, 149; and see also 230). Presumably, Rawls believes that constitutional consensus contributes to the forms of social unity and stability found in existing liberal societies, in which the requisite agreement can be found. He adds, moreover, as we saw above, that the smooth functioning of political institutions itself encourages "the cooperative virtues of political life," reasonableness, a sense of fairness, willingness to compromise, and "willingness to cooperate with others on political terms that everyone can publicly accept" (Rawls 1993, 163). If all this is true, then one must ask how constitutional consensus falls short.

Rawls holds that the two forms of consensus differ mainly in terms of breadth and depth.[30] In regard to the former, overlapping consensus encompasses a wider range of principles, basically more secure views of substantive rights, and more developed principles of economic distribution. Its superior depth lies in the fact that participants trace their views back to fundamental intuitive ideas rooted in the political culture. The principles that constitute a constitutional consensus are less firmly grounded.

Rawls believes that, as a result of these differences, overlapping consensus will be superior from a normative standpoint and also possess greater "stability." The essence of the former criticism has been discussed above. Even if we concede Rawls's belief that principles which provide more substantial economic protection are morally superior to those in a constitutional consensus (though, again, not everyone will accept this), he has not shown that the difference principle will be the focus of agreement throughout society, or that his method is the one best suited to develop principles that will be able to secure agreement.[31]

More important to Rawls are considerations of stability. We have seen that he faults the modus vivendi because, owing to its lack of moral depth, its participants will withdraw their allegiance in accordance with shifts in the balance of power. Rawls apparently believes that something similar holds in regard to consti-

---

[27] It bears mention that the general opinion of scholars is that a principle along these lines, rather than the difference principle, would actually be the choice of Rawls's representative individuals behind the veil of ignorance in *Theory of Justice*. As Barry says of Rawls's derivation of the difference principle, "no other aspect of the theory has met with such uniform rejection" (Barry 1989, 214; and for further references see p. 411 n1).

[28] Discussion of the economic principles to which convergence would give rise—if any (see below, footnote 31)—falls beyond the scope of this article, though it is apparent that the principles it would produce would differ substantially from Rawls's.

[29] See also Rawls 1996, xlvii–xlviii. In addition, Rawls's main account of the differences between the two kinds of consensus (1993, 144–68)

is less clear than it might be, because much of this discussion is drawn, with only slight revisions, from Rawls 1987, which was written before Baier's seminal article had brought the idea of constitutional consensus to Rawls's attention. Section 7, on the movement from constitutional consensus to overlapping consensus, is new, but the context into which it is inserted is not sensitive to the relationship between the two forms of consensus.

[30] He also discusses differences in specificity (Rawls 1993, 164–8), though these seem less significant (see the following note).

[31] Rawls notes the unlikelihood that there will be a set of principles of distribution upon which everyone will agree. Rather, it is "more realistic and more likely" that there will be a number of competing distributive principles, which will include justice as fairness (Rawls 1993, 164). If this is true, then the difference between overlapping consensus and constitutional consensus in regard to principles of economic distribution is narrowed, and Rawls's grounds for rejecting the former are also narrowed.

tutional consensus. Though he never discusses this matter directly, he apparently believes that it, too, is lacking in stability.[32]

In order to assess Rawls's argument, we must discuss the nature of stability and the distinctive way Rawls employs this concept. As encountered generally in discussions in political science and sociology, the term refers to attributes of a system of political institutions, namely, that it functions smoothly and is not plagued by large-scale, extralegal political opposition (see Lipset 1981, 30). For ease of reference, I will call this stability in the usual sense or political stability.

The idea of stability plays a significant role in Rawls's theory. For instance, he writes that *"all* differences" between *Political Liberalism* and *Theory of Justice* can be traced to problems in the view of stability put forth in the latter (1993, xv–xvi; emphasis added).[33] Thus, it is important to realize that Rawls discusses an unusual sense of the term, a psychological sense, which centers on the way in which the inhabitants of a regime hold their moral principles. In keeping with his view of moral psychology (Rawls 1993, 81–8), Rawls argues that people are more likely to act on their moral principles when they believe that their political system functions in accordance with them and that other citizens are also likely to behave in accordance with them (Rawls 1993, 86, 163; 1971, 138, 177, 454–5). It is because overlapping consensus has a superior ability to generate stability in this sense that we should prefer it to constitutional consensus.

There are several problems with this argument. Rawls is aware that considerations of stability (as he construes this) have gone largely undiscussed by previous theorists (Rawls 1993, xvii). I believe this neglect is justified. One can ask whether considerations of stability, in Rawls's sense, are sufficient to bear the weight he places on them.

It would be difficult for Rawls to argue that stability in his sense is necessary for the smooth functioning of liberal societies—for political stability. As we have seen, his criticism of modus vivendi centers on the claim that it will lack political stability. But a similar argument is not convincing in regard to regimes characterized by constitutional consensus. A striking fact of post–World War II political life is the obdurate stability of liberal democracies. This has been true in virtually all cases (with rare exceptions, such as Northern Ireland) for more than fifty years. Accordingly, if we assume that these regimes are characterized by constitutional consensus, which probably plays some role in their political stability, then it would be difficult for Rawls to argue that lack of stability in his sense undermines political stability. Moreover, Rawls's claim that stability in his sense contributes significantly to political stability is supported by little or no evidence and is not intuitively convincing. Rawls says that he is interested in a specific kind of political allegiance: "the bases of allegiance [a liberal conception of justice]

generates in virtue of the distinctive content of its principles" (Rawls 1993, 161 n26). In his latest discussion, Rawls refers to this as "stability for the right reasons" (1996, xxxix n5) In comparison to other factors that political observers since the time of Plato and Aristotle have identified as sources of political stability—for example, absence of economic, religious, racial, and ethnic differences—one wonders how great a role is played by specific aspects of the contents of liberal principles. Accordingly, though once again, overlapping consensus is not undesirable, Rawls has not made a strong case that its superior contribution to stability in his sense is sufficiently important to the functioning of a liberal regime to justify jettisoning a method of political construction that can generate constitutional consensus in favor of one that is otherwise lacking support but possibly able to generate overlapping consensus.

It is still open to Rawls to argue that stability in his sense is inherently valuable. Considerations along these lines are more important to him than political stability. Briefly, he claims the superiority of a regime with moral principles that encourage the cooperative virtues, thereby helping people fully develop their moral powers—in this case, specifically, the "reasonable." Rawls describes this as an ability "to understand, apply, and to act from the reasonable principles of justice that specify fair terms of social cooperation" (1993, 103–4) Yet, Rawls does not provide evidence for his crucial claim that the motivating force of moral principles with particular content is greater than that of other factors. Even if we concede that the factor he notes plays a role in influencing behavior, he does not address the question of how this factor interacts with others that also influence conduct, such as self-interest, religion, and national identification. In order for stability in his sense to be a central consideration in the representative individuals' deliberations, it must play a significant role in stimulating cooperative behavior, but this Rawls has not shown.

It is likely that the prominent role Rawls accords this factor is bound up with his conception of the person, as possessing the two moral powers. It appears that Rawls's postulation of the "reasonable" is less controversial than the aspects of his conception of the person discussed above. As we have seen, however, he presents little evidence that his particular conception of the person is deeply rooted in the public culture of liberal societies. Absent convincing arguments that inhabitants of liberal societies would accept this conception of the person, Rawls would find it difficult to demonstrate that a society in which moral principles are held in accordance with his notion of stability is significantly superior to one in which they are not.

In sum, even if we concede that a society with an overlapping consensus is preferable in various ways to one with only a constitutional consensus, Rawls does not make a convincing case that the latter is radically defective, and therefore it is necessary to move to overlapping consensus. The point bears repeating that, even if Rawls did make this case, it still would not demonstrate that his particular method is the one best

---

[32] See footnote 3.
[33] For stability's crucial role in derivations of principles of justice in Rawls's works, see Cohen 1989.

suited to develop an overlapping consensus. The possibility would remain that some other method is better suited to this task, and without the severe flaws in Rawls's overlapping consensus.

In conclusion, distinctive features of Rawls's method of constructivism are difficult to defend. The method employed by political liberalism should be the one most likely to generate principles that will satisfy the two conditions of being normatively robust and acceptable to diverse groups. Even if the first of these requirements takes clear precedence over the second, there do not appear to be strong grounds for concentrating on selected intuitive ideas and rigidly separating the stages of construction. A more advisable procedure would be to explore society's comprehensive views for areas of overlap and then work from these toward the best principles possible in the normative sense.

We have seen that Rawls pursues his constructivist strategy because he does not want political liberalism to be "political in the wrong way." One conclusion of this paper is that a proper method of political construction must be heavily political. In being worked up from existing comprehensive views, political liberalism must have a strong empirical basis. But we should note that the tasks of "political" liberalism do not exhaust the role of political philosophy. Though principles of justice upon which different groups are most likely to agree are closely related to those they already accept, another and more traditional endeavor of political philosophy is to defend what one can regard as true principles of justice (if only, perhaps, from the perspective of one's own comprehensive view) and to attempt to bring other members of society around to one's view. But this dual task of philosophical construction and persuasion should be distinguished from "political" liberalism. Different forms of construction that one can envision here are closely related to what transpires in the first stage of Rawls's constructivism. But because these are not hampered by the requirements of Rawls's second stage, they are able to proceed with less attention to unsettling aspects of the existing political world.

REFERENCES

Baier, Kurt. 1989. "Justice and the Aims of Political Philosophy." *Ethics* 99(July):771–90.

Barry, Brian. 1989. *Theories of Justice. A Treatise of Social Justice, Vol. I.* Berkeley: University of California Press.

Barry, Brian. 1995. Review of *Political Liberalism, Ethics* 105(July): 874–915.

Cohen, Joshua. 1989. "Democratic Equality." *Ethics* 99(July):727–51.

Cohen, Joshua. 1993. "Moral Pluralism and Political Consensus." In *The Idea of Democracy*, ed. David Copp, Jean Hampton, and John Roemer. Cambridge: Cambridge University Press.

Connolly, William. 1983. *The Terms of Political Discourse.* 2d ed. Princeton: Princeton University Press.

Easton, David. 1965. *A Systems Analysis of Political Life.* Chicago: University of Chicago Press.

Exdell, John. 1994. "Feminism, Fundamentalism, and Liberal Legitimacy." *Canadian Journal of Philosophy* 24(September):441–63.

Gallie, W. B. 1955–56. "Essentially Contested Concepts." *Proceedings of the Aristotelian Society* 56:167–98.

Galston, William. 1989. "Pluralism and Social Unity." *Ethics* 99(July):711–26.

Galston, William. 1992. *Liberal Purposes.* Cambridge: Cambridge University Press.

Hibbing, John, and Elizabeth Theiss-Morse. 1995. *Congress as Public Enemy.* Cambridge: Cambridge University Press.

Hunter, James. 1990. *Culture Wars: The Struggle to Define America.* New York: Basic Books.

Hunter James. 1994. *Before the Shooting Begins: Searching for Democracy in America's Culture War.* New York: Free Press.

Hunter, James, and Carl Bowman. 1996. *The State of Disunion: 1996 Survey of American Political Culture.* Ivy, VA: In Media Res Educational Foundation.

Jelen, Ted, and Clyde Wilcox. 1995. *Public Attitudes toward Church and State.* Armonk, NY: M. E. Sharpe.

Jones, Peter. 1995. "Review Article: Two Conceptions of Liberalism and Two Conceptions of Justice." *British Journal of Political Science* 25(October):515–50.

Klosko, George. 1993. "Rawls's 'Political' Philosophy and American Democracy." *American Political Science Review* 87(June):348–59.

Klosko, George. 1994. "Rawls's Argument from Political Stability." *Columbia Law Review* 94(October):1882–97.

Lipset, Seymour. 1981. *Political Man: The Social Bases of Politics.* Rev. ed. Baltimore, MD: Johns Hopkins University Press.

Miller, David. 1992. "Distributive Justice: What the People Think." *Ethics* 102(April):555–93.

Muller, Edward, Thomas Jukam, and Mitchell Seligson. 1982. "Diffuse Political Support and Antisystem Political Behavior: A Comparative Analysis." *American Journal of Political Science* 26(May): 240–64.

Rawls, John. 1958. "Justice as Fairness." *Philosophical Review*, 67(January):164–94.

Rawls, John. 1971. *A Theory of Justice.* Cambridge, MA: Harvard University Press.

Rawls, John. 1980. "Kantian Constructivism in Moral Theory." *Journal of Philosophy* 77(September):515–72.

Rawls, John. 1982. "The Basic Liberties and Their Priority." In *The Tanner Lectures on Human Values*, ed. Sterling MacMurrin. Cambridge: Cambridge University Press.

Rawls, John. 1985. "Justice as Fairness: Political Not Metaphysical." *Philosophy and Public Affairs* 14(Summer):223–51.

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

Rawls, John. 1988. "The Priority of Right and Ideas of the Good." *Philosophy and Public Affairs* 17(Fall):251–76.

Rawls, John. 1989. "The Domain of the Political and Overlapping Consensus." *New York University Law Review* 64(May):233–55.

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

Rawls, John. 1996. *Political Liberalism.* Paperback ed., with new Introduction. New York: Columbia University Press.

Wenar, Leif. 1995. "*Political Liberalism*: An Internal Critique." *Ethics* 106(October):32–67.

Wolff, Robert P. 1977. *Understanding Rawls.* Princeton: Princeton University Press.