At different stages of his career, John Rawls has attempted to justify his principles of justice through different methods. In his classic article, "Justice as Fairness," published in 1958, he relies on a version of the social contract argument. A Theory of Justice, published in 1971, presents a more elaborate contract argument, employing the now familiar device of representative individuals deliberating behind a "veil of ignorance" in the "original position" and important related notions, such as the doctrine of primary goods. Since the publication of A Theory of Justice, Rawls has moved towards a sociological account of his principles, which are represented as the core of a "political" conception of justice and which can constitute the heart of an "overlapping consensus" in contemporary liberal society. (These ideas are discussed below.) A developed account of this last position is presented in Rawls's recent work, Political Liberalism, aspects of which are examined in this Essay. It is striking that, as Rawls has moved between these justificatory models through a period of more than thirty-five years, there has been little change in the content of his moral view. Throughout this period, he has continued to defend his two principles of justice and the overall view he refers to as "justice as fairness" with relatively little alteration.

Two distinct sides to Rawls's sociological defense of his principles can be distinguished. First, he claims that justice as fairness is worked up from "certain fundamental ideas seen as implicit in the public political culture of a democratic society" (p. 15). The basic idea here is that Rawls believes his view is most "congenial to" or represents the most consistent...
elaboration of central aspects of democratic political culture. A second, increasingly important claim in Rawls's recent works is that a particular kind of moral/political view he calls "political liberalism" is necessary for the stability of liberal society, and that justice as fairness is such a view. The subject of stability receives considerable attention from Rawls. In *Political Liberalism* he says that the need to remedy defects in his account of stability in *A Theory of Justice* led to all revisions of his theory in the later work (p. xvii). In other respects, Rawls takes "the structure and content of [A] Theory [of Justice] to remain substantially the same" in the later work (p. xvi).

This Essay examines Rawls's conception of stability and its role in justifying his moral principles. Though Rawls places great weight on stability, his account of it is unsatisfying in important ways. In the Introduction to *Political Liberalism*, he identifies "the problem of political liberalism" as follows: "How is it possible that there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable though incompatible religious, philosophical, and moral doctrines?" (pp. xviii, 4). However, some aspects of Rawls's view of a stable society are unclear, while others are highly idiosyncratic. Part I of this Essay reviews Rawls's account of stability in *A Theory of Justice* and an ambiguity in his presentation in *Political Liberalism*. Part II examines the role of stability in establishing justice as fairness in the later work. In particular, this Part explores how considerations of stability would figure in the reasoning of the representative individuals in the original position in *Political Liberalism*, and whether these concerns would favor Rawls's preferred principles of justice. This Essay contends that Rawls faces a dilemma. Under those circumstances in which the representative individuals would view stability (what can be called "political stability") as an important concern, justice as fairness would make a relatively small contribution to it, and so considerations of stability would favor different principles. Under other circumstances, in which justice as fairness might do more to promote stability, the representative individuals would view stability itself as a minor concern and so would accord it little weight in their deliberations.

I. RAWLS'S CONCEPTION OF STABILITY

A. Stability in *A Theory of Justice*

In *A Theory of Justice*, Rawls justifies particular principles of justice by arguing that they would be chosen by suitably described representative individuals, deliberating behind a hypothetical veil of ignorance, in the original position. Because they want their chosen principles to endure, the representative individuals are concerned with the notion of stability. Rawls views stability as an attribute of moral principles. His concern with

this stems from certain "laws of moral psychology," which center on the 
generation of feelings of reciprocity: "[P]ersons tend to love, cherish, and 
support whatever affirms their own good."\textsuperscript{6} An individual brought up in 
a society in which everyone else adheres to moral rules, thereby confer-
ring great benefits on all members of society, including himself, will de-
develop a desire to comply as well. Rawls describes a set of moral principles 
as stable if it is able to generate attitudes that are sufficiently strong to 
outweigh other factors that would lead people to act unjustly:

\[W\]hen institutions are just (as defined by this conception), 
those taking part in these arrangements acquire the correspond-
ing sense of justice and desire to do their part in maintaining 
them. One conception of justice is more stable than another if 
the sense of justice that it tends to generate is stronger and 
more likely to override disruptive inclinations and if the institu-
tions it allows foster weaker impulses and temptations to act un-
justly. The stability of a conception depends upon a balance of 
motives: the sense of justice that it cultivates and the aims that it 
encourages must normally win out against propensities toward 
injustice.\textsuperscript{7}

Stable moral principles generate strong desires to comply throughout so-
ciety, thereby fostering general adherence to society's moral principles. 
Rawls describes such a society as "well-ordered." In such a society, "every-
one accepts and knows that the others accept the same principles of jus-
tice, and the basic social institutions satisfy and are known to satisfy these 
principles."\textsuperscript{8}

Rawls's discussion of stability immediately appears somewhat odd. As 
the term is generally used in discourse about politics, stability is an attri-
bute of political systems rather than of moral principles.\textsuperscript{9} While the con-
cept doubtless admits different interpretations, according to the most 
common usage, stable systems are characterized by law and order, 
smoothly functioning institutions, and similar features. Most obviously, 
such systems are without ruinous social conflict and other sources of dis-
ruption. A useful account of "political stability" is provided by Seymour 
Martin Lipset, according to whom a stable democracy is one that has en-
joyed "the uninterrupted continuation of political democracy since 
World War I and the absence over the past twenty-five years of a major 
political movement opposed to the democratic 'rules of the game.'"\textsuperscript{10}

\textsuperscript{6} A Theory of Justice, supra note 2, at 177.
\textsuperscript{7} Id. at 454.
\textsuperscript{8} Id.
\textsuperscript{9} A cursory search under the subject "political stability" of the University of Virginia 
Library system's periodical register (dating back to the early 1980s) indicates 204 titles. 
There are also 60 entries for books on this subject in the library index. Although I was not 
sure where to look for Rawls's concept of "stability," keyword searches of both data bases 
for "moral" and "stability" turned up no relevant entries.
This work was published originally in 1960. Klosko, supra note 5, at 349, also quotes this 
definition.
According to this common view, a "stable" democracy is characterized by smoothly functioning democratic institutions. By extension, other stable governments possess similar features.

Rawls's account of stability is importantly different. Rather than a characteristic of political systems, stability is an attribute of moral theories. A stable principle of justice is one that is able to give rise to certain attitudes. At one point, Rawls refers to this as the question of "psychological stability." For ease of reference, stability in this sense can be referred to as "moral stability."

Moral stability figures prominently in A Theory of Justice in the choice of moral principles in the original position. Though they are behind a veil of ignorance, the representative individuals are aware of "the general facts about human society," concerning such matters as political affairs, economic theory, and human psychology. Among such facts discussed by Rawls, those bearing on moral stability are prominent. When the representative individuals make their choice, a "strong point" in favor of a particular conception of justice is that it gives rise to the attitudes in question and so generates its own support. Because his principles evince greater concern for the value of each individual and the crucial primary good of self-respect, Rawls believes they will be more stable than utilitarian alternatives.

---

11. See A Theory of Justice, supra note 2, at 177.
12. See id. at 137. Exactly what these facts are is a matter of some controversy, while reliance on different sets of general facts could well lead the representative individuals to select different principles of justice. Rawls has been criticized for not adequately exploring the implications of possible disagreements along these lines. See, e.g., Richard W. Miller, Rawls and Marxism, in Reading Rawls, supra note 4, at 206. Because Rawls supplies the representative individuals with information relevant to their deliberations throughout A Theory of Justice, he is unable to avoid the implication that the information in question is his own understanding of these facts, that is, what he accepts as true about politics, society, psychology, and so on. For one area in which Rawls's understanding of an important subject matter is open to question, see infra Part II.A.
14. See id. at 177.
15. These considerations are closely related to Rawls's argument from the "strains of commitment." See id. at 176. Because the choice of moral principles in the original position is "final and made in perpetuity," id., the representative individuals must select principles that society will accept. Once again, because his principles are more deeply concerned with the well-being of each individual than are utilitarian alternatives, Rawls believes people would be better able to comply with them. Concerns of stability and strains of commitment are the main considerations Rawls advances in A Theory of Justice, Section 29, "Some Main Grounds for the Two Principles of Justice," which provides additional support for his three major arguments, presented in Section 26. For the appeal to stability, see id. at 177. Briefly, Rawls's three main arguments explain why the representative individuals should take a conservative approach to dealing with risk, which in turn leads them to select his principles. The three arguments turn on the following factors: the difficulty of estimating probabilities, diminishing marginal utilities above a certain threshold, and the grave risks associated with other outcomes. See id. at 154–55; see also Thomas Nagel, Rawls on Justice, in Reading Rawls, supra note 4, at 1, 10–15.
Rawls's overall reasoning in regard to stability can be surmised. He appears to believe that the representative individuals will be interested in moral stability, because this contributes to political stability, which is a necessary condition for acceptable lives. If this is actually Rawls's argument, then it depends on two assumptions. First, there must be significant differences between the contributions different moral principles make to political stability, and the contribution his principles would make must be significantly greater than that of other principles, especially utilitarian ones. Unless the difference between the respective contributions is substantial, the representative individuals would not appeal to this factor in their reasoning. Second, political stability must be an important consideration. If the representative individuals were not concerned with the threat of instability, the added contribution of Rawls's principles would not be a "strong point" in their deliberations.\textsuperscript{16}

B. Stability in Political Liberalism

Rawls attributes the major differences between \textit{Political Liberalism} and \textit{A Theory of Justice} to his realization of the irreducible pluralism of contemporary liberal societies. If widespread disagreements about basic moral and political questions are unavoidable, then the model of a well-ordered society presented in \textit{A Theory of Justice} is, as Rawls says, "unrealistic" (p. xvii).

In \textit{Political Liberalism} (and a series of articles leading up to it)\textsuperscript{17} Rawls attempts to develop what he calls a "political" conception of justice. In contrast to other moral doctrines, a political conception of justice is neither "general" in scope nor "comprehensive." By the former, Rawls means that a "political" view does not address a wide range of questions, but focuses on central political and social institutions. Unlike a comprehensive doctrine, which "covers all recognized values and virtues within one rather precisely articulated system" (p. 13), a political view attempts to stay on the philosophical surface and so avoids being dependent on controversial philosophical views. This allows a political view to be acceptable to people who subscribe to different comprehensive views. Whereas \textit{A Theory of Justice} does not distinguish between comprehensive

\textsuperscript{\textsuperscript{16}}To some extent this second assumption is complicated by an ambiguity in Rawls's understanding of stability. As he uses the term, a "stable" society is characterized not only by an absence of strife but also by the fact that institutions continue to function morally. See \textit{A Theory of Justice}, supra note 2, at 457-58. Thus it is not always clear with which sense of stability we are concerned. One respect in which the two senses of political stability are related is that, in a stable society, the just institutions put into effect at some previous time are what continue to function.

and political views, and so assumes that all members of society will understand Rawls's principles in the same way, *Political Liberalism* presents a toned-down account of a well-ordered society.

Because of the problem of pluralism, Rawls attempts to generate a set of principles that can be endorsed for different reasons by different people, each of whom views them from his or her own distinctive perspective. No longer attempting to produce moral unanimity, Rawls strives for principles that are able to give rise to an "overlapping consensus." He believes that different groups will be able to accept a set of common principles to adjudicate important political issues, though they disagree about wider philosophical questions, and accept the principles for different reasons (pp. 3-46, 133-72).

Though Rawls's concern with pluralism leads him to revise his theory, central features of *A Theory of Justice* survive in *Political Liberalism*. Importantly, Rawls retains both the concepts of the original position and the well-ordered society. He says that the original position still supports the same two principles of justice—though, since publication of *A Theory of Justice*, he has modified the first, the equal liberty principle, in response to important criticisms presented by H.L.A. Hart. In *Political Liberalism*, the principles are intended to support a well-ordered society—with the qualification noted. Rawls continues to endorse the moral psychology presented in *A Theory of Justice* (pp. 81-88, 140-44), and is still deeply concerned with the moral stability of principles of justice.

One difference between the works is that in *Political Liberalism*, Rawls is more deeply concerned with political stability. His greater attention to this problem is apparent in the two fundamental questions *Political Liberalism* addresses. The first, a normative question, is similar to the moral inquiry that dominates *A Theory of Justice*. As stated in Lecture I of *Political Liberalism*, this is as follows: "[W]hat is the most appropriate conception of justice for specifying the fair terms of social cooperation between citizens regarded as free and equal . . . ?" (p. 3). The second question is initially described as "that of toleration understood in a general way" (p. 3). Restated more fully, the second fundamental question is "how citizens, who remain deeply divided on religious, philosophical, and moral doctrines, can still maintain a just and stable democratic society" (p. 10). Because it is used in reference to democratic society, the word "stable" in this last quotation obviously refers to political stability. Putting the two fundamental questions together, Rawls wishes to identify the best possible principles of justice in purely normative terms that will also promote political stability. These two aims are evident in his statement of "the problem of political liberalism" (p. xviii). But though Rawls now believes political stability is a fundamental concern, it still appears to be less im-


19. Throughout *Political Liberalism*, Rawls more frequently uses this term—as in *A Theory of Justice*—in regard to the stability of moral principles (pp. 78, 140 n.7, 148).
important than the normative appropriateness of principles—as indicated by the first fundamental question of political liberalism.20

In regard to the crucial question of how the stability of moral principles promotes political stability, Rawls's view in *Political Liberalism* closely resembles that in *A Theory of Justice*. His opinion in this regard is central to his important idea of the overlapping consensus.21 Rawls believes Western liberalism emerged from a cauldron of conflict between different groups during its formative centuries. Of special importance is the origin of religious toleration, which arose out of European religious wars in the sixteenth and seventeenth centuries. The fact of ineradicable pluralism led to "discovery of a new social possibility: the possibility of a reasonably harmonious and stable pluralist society" (p. xxv). Building on this example, Rawls argues that the task of political philosophy is set in motion by "deep political conflicts." It is utopian not to believe such conflicts exist. They generate the requirement that political philosophy bridge significant differences: "We turn to political philosophy when our shared political understandings . . . break down" (p. 44). It is notable that to illustrate such conflicts Rawls uses the example of disagreements between the American North and South over slavery (p. 45). Such cases demonstrate the importance of finding means for groups to coexist in peaceful, stable societies.

But it is not enough for Rawls that groups merely coexist. He distinguishes between what he calls a modus vivendi and an overlapping consensus. The modus vivendi is conceived on the model of a truce. Groups with different views agree to live and let live, but with the tacit proviso that their agreement will last only as long as the balance of power between them remains the same. Such an agreement is morally superficial; if the balance of power shifts, the dominant group is likely to attempt to renew conflict with its now weaker opponents (pp. 147–48). An example of a modus vivendi is the relationship in the sixteenth century between Catholics and Protestants, who came to endorse the principle of toleration out of necessity rather than conviction (p. 148).

In contrast to a modus vivendi, an overlapping consensus is morally stable in Rawls's sense. Its adherents support it for reasons that go beyond convenience. Toleration is rooted in their comprehensive moral views—though supported differently by the views of different groups—and fostered by the sense of justice that develops in a well-ordered society (pp. 140–50). A significant advantage of an overlapping consensus is its greater political stability. Groups are likely to adhere to it even if the balance of power shifts in such a way that one group would be able to take advantage of others (pp. 147–49).

20. An additional, strong indication that Rawls values normative superiority over political stability is the order of the two stages of the representative individuals' deliberations. See infra text accompanying notes 25–26.

21. See discussion infra.
If the representative individuals accept Rawls's assumptions about moral psychology and the relationships between different groups, they may well choose his principles of justice over possible alternatives. As Rawls says in *A Theory of Justice*: "Other things equal, the persons in the original position will adopt the more stable scheme of principles."\(^\text{22}\) Other things being equal, they would have little reason not to. But for their reasoning to be sound, Rawls's principles must make a real difference for political stability, and maintaining political stability must be a real concern.

Even if these assumptions are granted, however, Rawls's argument still confronts a possible conflict. The two fundamental questions of political liberalism indicate that the representative individuals must keep two considerations in mind: they must choose principles that are normatively superior, and that contribute most to a stable society.\(^\text{23}\) Rawls does not pay much attention to the possibility that these goals might diverge, apparently because of an additional assumption that the best moral principles will contribute most to political stability.\(^\text{24}\) This assumption rests on his view of the psychological qualities that proper moral principles generate. Accordingly, Rawls notes that if his moral psychology is incorrect, "there is a serious problem with justice as fairness" as he has presented it (p. 252).

Although *A Theory of Justice* recognizes that political stability influences the representative individuals' deliberations, Rawls's increased attention to problems of pluralism in *Political Liberalism* makes concerns of stability more pressing. However, though Rawls is aware that his defense of his principles of justice rests heavily on his moral psychology, he does not adequately address the concept of political stability and considerations that bear directly on it.

II. THE REPRESENTATIVE INDIVIDUALS' DELIBERATIONS CONCERNING POLITICAL STABILITY

This Part examines the role political stability should play in the representative individuals' deliberations. Though Rawls's increased attention to political stability in *Political Liberalism* is apparent in particular examples he uses, especially the wars of religion and other social conflicts, he does not explain exactly how stability affects the representative individuals' deliberations. Their reasoning, however, can be reconstructed.

Briefly, according to Rawls, the representative individuals deliberate in two stages. In the first, they begin with certain ideas Rawls views as fundamental to liberal political culture. These are the following concep-

---

24. Rawls argues that, other things being equal, the representative individuals will choose the most stable principle, thereby identifying it as best. But this argument rests on his assumption that a given principle will be most stable because of important normative characteristics. See infra note 26.
tions: first, society as a fair system of social cooperation; second, citizens as free and equal persons; and third, a well-ordered society, "effectively regulated by a political conception of justice" (p. 14). Rawls describes these ideas as "implicit in the public political culture of a democratic society" (p. 13). This important contention receives no detailed discussion, and Rawls does not explain exactly what he means by this or other similarly obscure formulations. But it appears his position is that these ideas are central to liberal political thought as a whole. Whatever their other differences, participants in diverse liberal traditions subscribe to these ideas, which can therefore be used as starting points for developing moral principles that otherwise disparate groups will accept.

Deliberating behind the veil of ignorance, the representative individuals are, of course, unaware of their particular attributes, such as economic status, age, sex, religion, race, and level of natural abilities. At the first stage, they are also unaware of their comprehensive moral views (pp. 24, 141–42). They choose suitable principles of justice as a "freestanding view" by working out the implications of the fundamental intuitive ideas (pp. 10, 12, 64, 144–45). Rawls fears that if principles were developed from the comprehensive views existing in society, principles of justice would be "political in the wrong way" (pp. 39–40, 142). A compromise between comprehensive views might resemble a modus vivendi, which is unstable in the long term.

After generating their principles in the first stage, the representative individuals attempt to determine whether these can support an overlapping consensus between diverse groups, and thus promote long-term stability. That Rawls regards political stability as a central concern is clear from this second stage, which is absent from A Theory of Justice. Because success at this stage is necessary (pp. 65–66, 78), initial selection of principles is provisional. Some tinkering might be needed to make sure principles are congenial to society (pp. 65–66). But although Rawls clearly recognizes considerations of political stability, the fact that he relegates them to this confirmatory stage indicates that they are secondary to concerns of justice.

A. The "Political Threshold"

Though Rawls recognizes the significance of political stability, his discussion is inadequate in various ways. To begin with, he never explains exactly what he means by political stability, or such related notions as "an enduring and secure democratic regime" (p. 38). This omission is significant because these concepts admit different interpretations. Rawls believes in the "burdens of reason," the fact that people legitimately, in-

25. See Klosko, supra note 5, at 350–51.
26. It is also likely that Rawls proceeds in this way because he believes success in the second stage will follow naturally after success in the first. Because of the laws of moral psychology, principles that have the most desirable normative features will also be the most stable morally and thus will do most to promote political stability.
deed, inevitably come to different conclusions in moral matters (pp. 54–58). Accordingly, they will subscribe to different views about the nature of a stable society, which is a complex, normative concept. It seems that Rawls pays little attention to considerations of political stability because he closely associates this notion with moral stability, and he believes that the latter will promote the former. In *Political Liberalism*, he suggests that a stable society is based on moral stability, without exploring other factors (pp. 140–44, 160–63).

A view of political stability more conventional than Rawls’s will recognize this quality’s specificity. It is central to conventional views of stability that particular factors promote stability in different ways in different societies. Realizing this could lead the representative individuals to endorse different moral principles for different societies. Social scientists traditionally recognize that factors affecting political stability vary with societies. Montesquieu provides a classic statement of this view in *The Spirit of Laws*. He argues that each country has a “general spirit” (esprit général), comprised of its climate, geological features, economic characteristics, religion, customs, etc., and that, in order to function properly, specific laws must be in accord with this. Because this is the dominant view among social scientists, it will be accepted by the representative individuals. Rawls provides no reason for setting this view aside in favor of his apparent belief that a single set of moral principles will most effectively promote stability in all liberal societies.

Rawls’s treatment of stability is one respect in which he deviates from his general practice of working within developed scholarly traditions. In many ways, Rawls views himself as a traditional scholar. For example, he makes frequent references to the history of moral philosophy and has doubts about moral conceptions that are not supported by a tradition of adherents (pp. 59–60). Rawls believes the problem of stability is “fundamental to political philosophy,” and is surprised it “has played very little role in the history of moral philosophy” (p. xvii). The reason for this is surely the oddness of his conception of stability. He all but ignores the problem of stability as it has been discussed by political philosophers and social scientists for centuries. His entire treatment of political stability is hampered by his failure to examine different factors that contribute to, or weaken, this. His emphasis on moral stability above all other factors

---

27. Rawls also explores more traditional, political conceptions of stability (pp. 38, 134).

28. See Montesquieu, *The Spirit of Laws*, bks. 1–8, 19, at 289–90 (David W. Carrithers ed. & Thomas Nugent trans., 1977) (1748). Detailed discussion of the traditional view of political stability is beyond the scope of this Essay. However, the prevalence of a view akin to Montesquieu’s is evident in the widespread opinion that he is a founder of modern social science. See, e.g., Raymond Aron, *Main Currents in Sociological Thought* 13–30 (Richard Howard & Helen Weaver trans., 1965); Louis Althusser, *Politics and History* (Ben Brewster trans., 1972); Emile Durkheim, Montesquieu and Rousseau: Forerunners of Sociology 61 (Ralph Mannheim trans., 1970).

29. See also *A Theory of Justice*, supra note 2, at 122–26, 581.
would strike most political sociologists as, at best, an unusual claim, lacking either empirical or philosophical support.

If the representative individuals are more attuned to the range of factors that affect political stability, they will take these into account in their deliberations. In particular, knowing that causes of stability vary, the representative individuals are likely to require detailed information about the particular societies they consider and could recommend different moral principles for different societies. This information would not violate the veil of ignorance, since they need not know to which particular societies they belong, to say nothing of their own particular attributes. Confirmation of the representative individuals' license to provide different principles for different societies is the fact that Rawls has them derive "special" and "general" conceptions of justice as fairness for different kinds of societies in *A Theory of Justice.*

Rawls distinguishes special and general conceptions of justice as fairness because he believes a society must meet a certain threshold of economic well-being before the two principles of justice are appropriate for it. Though Rawls nowhere pursues this implication, a similar distinction could be invoked in regard to matters of political stability. Because of their great concern with this subject, the representative individuals would do well to consider a related threshold bearing on the degree of conflict between groups and the resulting degree of social harmony. A basic distinction they could invoke is between what can be referred to as hostile and peaceful pluralist societies. If a society is so lacking in social concord that enjoyment of liberty is not possible, the representative individuals might be willing to curtail individual rights until the requisite level of security is attained. Their reasoning here would be analogous to what Rawls presents for the general conception of justice as fairness. Rawls argues that, when a society's level of economic development does not permit full enjoyment of rights, the representative individuals will decide to limit these to the extent necessary "to raise the level of civilization so that in due course these freedoms can be enjoyed." By appealing to a similar threshold in regard to political stability, the representative individuals would see the need to limit certain rights if this is necessary to create conditions under which full rights will be possible. For ease of reference, this Essay will refer to this threshold as the "political threshold," to distinguish it from the "economic threshold" to which Rawls appeals in distinguishing the general and special conceptions of justice as fairness.

Hostile pluralist societies fall below the political threshold. Violent conflict between groups tears these societies apart, even though they remain democratic. Examples include Northern Ireland, with its conflict.
between Protestants and Catholics, and India, with its array of religious conflicts. This list could be expanded to include non-democratic societies, or various societies in which democratic institutions are only beginning to take hold. Confining the category to democracies still allows it to include Rawls's example of the United States prior to the Civil War, though his example of Europe in the age of religious wars is ruled out. The most notable peaceful pluralist societies are the industrial democracies since World War II, all of which have had relatively little conflict between groups, though there has been at least some in most countries. It seems evident that these societies surpass the political threshold and so possess the amount of political stability necessary for full political rights, an approximation of which their members presently enjoy.

Once the political threshold is introduced and peaceful and hostile societies distinguished, Rawls's interest in promoting political stability raises a dilemma. In hostile societies, in which these concerns are most salient, the representative individuals are unlikely to choose Rawls's principles because others would make a more significant contribution. In peaceful societies, on the other hand, considerations of moral stability might favor Rawls's principles. However, such societies are already quite stable, and so marginal differences in moral stability between Rawls's principles and others would carry little weight in the representative individuals' deliberations.

B. Hostile Pluralist Societies

The representative individuals' deliberations concerning societies that fall below the political threshold should proceed along the following lines. To the extent they are influenced by considerations of stability, the representative individuals will probably choose principles for these societies with less moral content than Rawls's principles. Stability is only one concern among many, but Rawls recognizes that stable social and political systems are necessary conditions for the enjoyment of rights and liberties. To the extent they consider the need to create stability, the representative individuals could well choose principles that center on procedures to resolve disputes in ways contending groups would find acceptable. Insofar as they are concerned with stability, the representative individuals will recognize suitable procedural mechanisms as necessary prerequisites for other concerns. Something along these lines is chosen by the contending inhabitants of John Locke's state of nature, and is

33. Though these necessary conditions receive little direct discussion in *A Theory of Justice*, they are clearly assumed. See, e.g., id. at 212-13, 380. Rawls's increased emphasis on avoiding ruinous political conflicts in *Political Liberalism* (pp. xxiv, 44-45, 147-48) makes the need for a stable social and political system more evident in this work.

34. Locke argues that individuals in the state of nature will leave this condition by ceding certain of their rights to a neutral government that will adjudicate their disputes. See John Locke, Of Civil Government Second Treatise 78-80, 102-04 (Henry Regnery Co. 1955) (1689).
similar to the "constitutional consensus" found in existing liberal societies.

It is important to note that though principles of justice centering on these features protect individual rights, they fall short of the "priority of liberty" that is central to Rawls's principles. The need to make trade-offs between liberty and security is a familiar theme in political theory. Briefly, as individuals have greater reason to be concerned for their safety, they may be more willing to sacrifice some of their liberty for increased security. Under conditions of social unrest, they might consider relaxing various protections, such as the rights of suspected criminals. In extreme circumstances, they might welcome martial law. Because trade-offs can be necessary, the particular conception of rights expressed in Rawls's principles is more suited to some societies than others. For instance, in A Theory of Justice, Rawls argues that self-respect is the most important primary good. But clearly, in conditions of widespread unrest, self-respect would take a back seat to security. Similarly, the need for security—and so means to adjudicate between contending groups—would take precedence over expansive rights to free speech or association. Because of their emphasis on self-respect and the special requirements of the least advantaged members of society, Rawls's principles of justice are best suited to societies lacking significant social conflict and would be less likely than other principles to promote political stability under conditions in which political stability is a real concern.

Rawls anticipates an objection along these lines in Political Liberalism. He argues that his principles and the overlapping consensus to which they give rise are not utopian. Over time, with favorable circumstances, a modus vivendi can develop into an overlapping consensus as people become accustomed to increasing degrees of social harmony and their

35. See Kurt Baier, Justice and the Aims of Political Philosophy, 99 Ethics 775 (1989). Baier describes this "constitutional consensus" as a consensus "on the 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." Id. at 775. Baier argues that Rawls's theoretical needs would be better served by such an agreement than by a full-fledged overlapping consensus. Rawls responds to Baier's view in Political Liberalism (pp. 158-68). For further discussion of "constitutional consensus," see Klosko, supra note 5, at 355-56.


38. See A Theory of Justice, supra note 2, at 440-46.

39. Formally, the first principle of justice, as discussed in Political Liberalism's "The Basic Liberties and their Priority," might be able to accommodate severely restricted rights because it requires that the overall system through which rights are secured provide the most extensive rights possible under the circumstances (pp. 331-40). However, highly curtailed rights such as speech and association, even if formally compatible with the first principle, would significantly depart from the overall tenor of justice as fairness (or the special conception of it Rawls defends), as the general conception departs from the special conception.
moral psychologies evolve accordingly (pp. 140–50, 158–68). Because of the potential for greater stability to develop, Rawls believes the representative individuals would still choose his principles (pp. 158–68). However, he does not consider the possibility that preferable moral principles might vary with degrees of political stability. Even if the representative individuals would eventually support his principles, they might still choose a modus vivendi for the insecure conditions out of which stable societies emerge. A constitutional consensus would be especially attractive under conditions of strife, because it “may be sufficient for less demanding purposes and far easier to obtain” than a full overlapping consensus (p. 149). Rawls believes an overlapping consensus based on justice as fairness “would complete and extend” the development of liberal thought that began three centuries ago with acceptance of religious toleration (p. 154). But he does not draw the important implication that societies with different degrees of stability require different principles. Though Rawls repeatedly notes that his principles are intended for societies that possess the “reasonably favorable conditions that make democracy possible” (pp. 146, 155–56, 209), he does not discuss the possibility that this condition limits his principles’ applicability to societies that are basically harmonious.

C. Peaceful Pluralist Societies

It is possible that Rawls’s position could be salvaged for societies that surpass the political threshold and thus are without significant instability. Were he explicitly to limit his concern to such societies, Rawls could avoid having to demonstrate that his principles are the most appropriate for societies with significant conflict. Because most industrial democracies are generally harmonious, his achievement would still be important. In this case, however, the problem is how an absence of instability would influence the representative individuals’ deliberations.

Rawls’s understanding of modern democratic societies appears to be hampered by common misconceptions. He apparently endorses a range of views that are frequently referred to as “consensus theory.” As described by James Wright, this view holds that democratic political arrangements are inherently fragile; they “naturally tend toward collapse in the absence of any countervailing mechanism.”40 Because of the weakness of liberal institutions, proponents of consensus theory believe it is important to promote democratic values throughout society, in order to counteract destabilizing tendencies. Because the representative individuals share

40. James D. Wright, Political Disaffection, in 4 The Handbook of Political Behavior 12 (Samuel L. Long ed., 1981). Though Rawls does not explicitly endorse consensus theory, his great concern with stability is difficult to explain on any other understanding of democratic societies. It seems that Rawls makes similar, related mistakes in his suppositions about the political culture of liberal societies. See Klosko, supra note 5, at 355–56. In these respects, Baier’s description of Rawls’s arguments as “armchair sociology” seems appropriate. See Baier, supra note 35, at 783.
Rawls's adherence to consensus theory, they are deeply concerned with preserving the stability of liberal societies, and thus favor moral principles that support this.

The facts of political sociology over the past half century, however, sharply contradict Rawls's assumptions about democratic regimes. Brief reflection will reveal that, virtually without exception, the industrial democracies have been remarkably stable for almost fifty years. To quote Wright once again: "The notion that democracies are somehow inherently unstable, that they naturally tend toward collapse, is very difficult to square with this and most other aspects of world political history since World War II." He continues: "In the modern world, democracies prove not to be inherently unstable, as much theory suggests, but rather obdurately stable under almost all circumstances." If Wright is correct, then the representative individuals will not be much concerned with political stability, and considerations of moral stability will interest them even less. Their choice of moral principles will be based almost entirely on other factors.

Discussion here is complicated by the fact that political stability is, again, a complex, controversial notion. A mere absence of conflict (i.e., the essence of what is generally viewed as stability) is not necessarily valuable, unless it is accompanied by desirable moral qualities. Though the liberal democracies have been stable, they have been beset with discrimination against various minority groups and other significant forms of injustice. Unfortunately, in many countries, such phenomena appear to have increased in recent years, which makes bridging the differences between contending groups a crucial concern. Rawls's many contributions to our understanding of these matters in Political Liberalism significantly advance current debates. In their deliberations, the representative individuals would doubtless be concerned with the unequal treatment of certain groups and would wish to develop an overlapping consensus between their society's diverse comprehensive views. However, in choosing appropriate principles, they would pay little attention to concerns of political stability. The unequal treatment of certain groups—generally minority groups—is distinct from concerns of maintaining public order. Because targets of injustice are often small minorities—and often powerless as well—the representative individuals can view their persecution as compatible with the preservation of political stability. To proponents of liberal values, injustice and discrimination are, of course, objectionable and contrary to their basic moral beliefs. But to return to Rawls's two concerns of political liberalism, these matters bear more directly on fair terms of cooperation between groups than on preserving a stable society. It seems that in Political Liberalism—and to a lesser extent in A Theory of Justice—Rawls runs together these different concerns and rests far too much of his case for his moral principles on stability.

41. Wright, supra note 40, at 67.
Conclusion

Examination of the role that considerations of stability should play in the representative individuals' deliberations in Political Liberalism suggests that these concerns undermine Rawls's defense of justice as fairness. Rawls notes that political liberalism is a family of doctrines that is capable of contributing to an overlapping consensus in society. Considerations of political stability may well favor different moral principles for different societies. For hostile or unstable societies, the representative individuals might support a variant of political liberalism with a conception of rights less demanding than Rawls's own. For peaceful or stable societies, on the other hand, they might favor Rawls's principles, though not for reasons of moral stability. Considerations of stability, once again, are only one topic among many that the representative individuals would consider. But there is a possible problem in that, beyond a certain point, arguments for Rawls's principles on grounds other than stability run the risk of ceasing to be compatible with conflicting comprehensive views, and thus ceasing to be "political" liberalism. These suggestions clearly require a great deal more discussion. As things stand, however, it appears that Rawls's support of justice as fairness on grounds of its superior moral stability requires substantial revision. This is indeed "a serious problem with justice as fairness" as Rawls has presented it (p. 252)—though rather more for reasons of political sociology than moral psychology.