

## RAWLS'S LIBERAL PRINCIPLE OF LEGITIMACY

EDWARD SONG

The problem of legitimacy has received a great deal of attention in recent years, but very little of it has been devoted toward examining John Rawls's liberal principle of legitimacy as a self-standing theory. This principle plays an important role in Rawls's reinterpretation of justice as fairness as a *political* theory of justice. Part of the reason for this lack of attention is the fact that Rawls himself says little to develop or explain the principle, and it is often lost in the controversial debates about the necessity of his turn to "political liberalism." Nevertheless, it suggests a highly original way of thinking about state legitimacy that offers an unusual alternative to the dominant theories that are currently available. In this paper, I will attempt to develop a sketch of what the general contours of such an account might look like. The goal is not to read the mind of Rawls, or offer a precise textual exegesis, but rather to develop an interpretation of the principle that is rooted in Rawls's brief remarks about it and the core features of political liberalism. At the heart of this account is the idea that the legitimacy of the state resides not in the consent of the governed, nor in the state's conformity with the appropriate principles of justice, but rather in citizens' endorsement of the state and its underlying constitutional principles on the basis of their own comprehensive conceptions of justice. Such an account offers a stable middle ground between popular alternatives, and has a way of not just solving, but dissolving, the problem of legitimacy.

### I. STABILITY AND THE PROBLEM OF CONGRUENCE

One challenge in interpreting Rawls's liberal principle of legitimacy is born from the fact that its genealogy is complex, arising out of his earlier discussion of political stability in *A Theory of Justice*. This is contentious terrain since the problem of political stability is the source of Rawls's dissatisfaction with justice

as fairness as it was originally articulated, and gives rise to the controversial modifications discussed in *Political Liberalism*. In *Theory*, Rawls claims that the primary role of principles of justice is to “specify a system of cooperation designed to advance the good of those taking part in it.”<sup>1</sup> In addition, however, Rawls claims that there are “other fundamental social problems” that are “connected with that of justice.”<sup>2</sup> Among these is the problem of political stability. While this issue is conceptually distinct from an inquiry into the nature of justice, he suggests that “we cannot, in general, assess a conception of justice by its distributive role alone. We must take into account its wider connections; for even though justice has a certain priority, being the most important virtue of institutions, it is still true that, other things equal, one conception of justice is preferable to another when its broader consequences are more desirable.”<sup>3</sup> More specifically, he adds that “If a conception of justice is unable to generate its own support, or lacks stability, this fact must not be overlooked. For then a different conception of justice might be preferred.”<sup>4</sup> Thus, the concept of stability is *external* to the concept of justice, yet it operates as a kind of feasibility test that a theory of justice must pass. For if that conception of justice failed to engender political stability, then it ought to be abandoned in favor of one that would.

Rawls thinks that a conception of justice would be stable when the institutions it recommends are “more or less regularly complied with and its basic rules willingly acted upon; and when infractions occur, stabilizing forces should exist that prevent further violations and tend to restore the arrangement.”<sup>5</sup> More specifically, Rawls’s inquiry into stability is directed toward two distinct but related issues: the capacity of a conception of justice to encourage a “sense of justice” in its citizens, and the wider and more vexed problem of the “congruence” between goodness and justice.

A sense of justice is the “normally effective desire to apply and to act upon the principles of justice, at least to a certain minimal degree.”<sup>6</sup> Without such an assumption, the original position contractors would be unwilling to commit themselves to any principles of justice because such a contract is attractive only when one has assurances that all others can and will make a good faith effort to abide by the terms of the agreement. Rawls suggests that justice as fairness tends to encourage such a sense of justice since all persons are guaranteed maximal freedom and adequate social provisions. Justice as fairness has the tendency to

---

<sup>1</sup> John Rawls, *A Theory of Justice* (Cambridge, MA: Belknap Press, 1971), 4.

<sup>2</sup> *Ibid.*: 6.

<sup>3</sup> *Ibid.*

<sup>4</sup> *Ibid.*: 145.

<sup>5</sup> *Ibid.*: 6.

<sup>6</sup> *Ibid.*: 505.

encourage allegiance to the institutions that it recommends, since these institutions reciprocally promote everyone's good.<sup>7</sup>

That a conception of justice encourages a sense of justice, however, is not itself sufficient to ensure its stability. For even if institutions ordered according to justice as fairness in fact encourage a sense of justice, the stability of such a polity might still be threatened by what Rawls calls the problem of the "congruence" of goodness and justice. The deeper problem is, as Rawls says, "whether being a good person is a good thing for that person."<sup>8</sup> In other words, is justice part of the human good? Wouldn't it be better for us, from the perspective of our own individual well-being, if we were able to receive all of the benefits of living in a cooperative social scheme, but avoid all of the costs, by free riding on the sacrifices of others?

Rawls argues that in a well-ordered society structured according to justice as fairness, justice and goodness are in fact congruent—a conclusion that is noteworthy in its own right apart from its implications for the stability of his theory.<sup>9</sup> His argument for this conclusion is rooted in three basic claims. First, the person for whom the vantage points of justice and goodness are never reconciled, is constantly at odds with his own sense of justice and must confront the psychological costs "of pretending to have certain moral sentiments, all the while being ready to act as a free rider whenever the opportunity arises to further his personal interests."<sup>10</sup> Second, Rawls appeals to what he calls "the Aristotelian Principle" to suggest that the sense of justice makes possible a form of human sociability that is a great good to individuals. As Rawls says, "human beings have in fact shared final ends and they value their common institutions and activities as good in themselves. We need one another as partners in ways of life that are engaged in for their own sake, and the successes and enjoyments of others are necessary for and complimentary to our own good."<sup>11</sup>

Finally, Rawls appeals to the "Kantian interpretation" of justice as fairness, suggesting that the desire to abide by the principles of justice is in accord with the deepest aspects of our natures as autonomous creatures. Rawls understands the original position to offer a kind of procedural interpretation of Kant's moral philosophy. The original position argument models the Kantian ideal of autonomy

<sup>7</sup> Ibid: 453–512.

<sup>8</sup> Ibid: 397.

<sup>9</sup> Samuel Freeman notes that "Rawls has said (in conversation) that he thinks the congruence argument was one of the most original contributions he made in *A Theory of Justice* and that he is puzzled why it did not attract more comment." "Congruence and the Good of Justice," *The Cambridge Companion to Rawls*, ed. Samuel Freeman (Cambridge: Cambridge UP, 2003), 308, n. 2.

<sup>10</sup> Rawls (1971): 570.

<sup>11</sup> Ibid: 522–3.

since the two principles of justice are those principles that would be adopted by rational persons choosing from a position of freedom and equality. The veil of ignorance deprives persons of the knowledge of their contingent particularities, facts which, if they were the sole motivation, would result in heteronomous action. As Rawls says, “properly understood, then, the desire to act justly derives in part from the desire to express most fully what we are or can be, namely free and equal rational beings with a liberty to choose.”<sup>12</sup>

These three considerations are the chief reasons in favor of the congruence of the demands of justice with an individual’s own good. In short, they all suggest that persons who violate their sense of justice will be acting against various deep aspects of their natures. Rawls takes these reasons to be decisive, especially when contrasted with the plausibility of congruence for other prominent accounts of justice, and as a result, there is no need to reconsider justice as fairness and the two principles that it recommends. A society ordered according to justice as fairness is as stable as one could hope.

Rawls, of course, later comes to see these arguments for the stability of justice as fairness to be critically flawed, and it is these worries that lead to the reinterpretation of the theory as a political theory of justice. Briefly, from the perspective of *Political Liberalism*, the problem with justice as fairness as it was originally articulated is that it is a “comprehensive” moral theory, one that “applies to a wide range of subjects (in the limit to all subjects); it is comprehensive when it includes conceptions of what is of value in human life, as well as ideals of personal virtue and character, that are to inform much of our non-political conduct.”<sup>13</sup> In short, it is a full-blown philosophical account of morality and the human good. As such it would fail to insure sufficient stability as a charter for political institutions in the kind of diverse settings that we find in contemporary societies. For such societies are inevitably characterized “by a pluralism of incompatible yet reasonable comprehensive doctrines. No one of these doctrines is affirmed by citizens generally. Nor should one expect in the foreseeable future one of them, or some other reasonable doctrine, will ever be affirmed by all, or nearly all, citizens.”<sup>14</sup> The problem with justice as fairness as it was articulated in *Theory* is that it is too partisan a comprehensive philosophical theory to serve as a shared justification for a liberal theory of justice in pluralistic societies.

Rawls addresses this problem of political instability by redescribing justice as fairness as a *political* conception of justice that could be endorsed by almost anyone regardless of their own comprehensive philosophical commitments. In

---

<sup>12</sup> Ibid: 256.

<sup>13</sup> John Rawls, *Political Liberalism* (New York: Columbia UP, 1996), 175.

<sup>14</sup> Ibid: xvi.

contrast to comprehensive theories, political theories are defined by their limited scope, applying only to the sphere of political and social institutions, and saying nothing about a wider range of moral issues. Justice as fairness is a moral conception, but as Rawls says, "it is a moral conception worked out for a specific kind of subject, namely for political, social and economic institutions."<sup>15</sup> Understood as a political conception, justice as fairness can serve as the focal point of an "overlapping consensus of reasonable comprehensive views." Thus, the content of the theory has not changed, but Rawls now imagines it to arise out of the ideas latent in the public political culture of liberal democratic society. It is a kind of freestanding module, dependent on no particular comprehensive doctrine, and acceptable by all reasonable doctrines. As such, it can serve as a unifying charter for pluralistic liberal societies.

The modifications of political liberalism, however, are controversial, and many think that they are unnecessary. The most forceful of these critics is Brian Barry, who argues that the theory as it was originally articulated is stable, even from the perspective of Rawls's later work.<sup>16</sup> Indeed, Barry argues that the later Rawls attributes to the Rawls of *A Theory of Justice* errors that he did not in fact commit, and that the best thing to do is to simply forget that the later book had ever been written.

Barry contends that the modifications of political liberalism are unnecessary because "it is hard to see why *A Theory of Justice* should not be said to contain a political conception of justice."<sup>17</sup> He argues that, "the whole point of *A Theory of Justice*, was that it left people to form, revise and pursue their own conceptions of the good. (That is why agents in the original position did not know their conception of the good: they would have no incentive for building into the principles of justice a preference for any particular conception of the good.)"<sup>18</sup> No person, regardless of her conception of the good, has reason to resist justice as fairness precisely because the theory gives people the freedom to live their lives as they see fit.

If comprehensive elements do enter the theory, Barry argues that they do so only because of the congruence argument. For there, in order to show that the inclination to live within the bounds prescribed by justice is compatible with and conduces to the human good, Rawls must appeal to "the Kantian Interpretation," a theoretical component that is nothing, if not comprehensive. The congruence argument, therefore, infects the rest of the otherwise political theory with controversial comprehensive elements. If the congruence argument was unnecessary,

---

<sup>15</sup> Ibid: 11.

<sup>16</sup> Brian Barry, "John Rawls and the Search for Stability," *Ethics* 105 (July 1995): 874–915.

<sup>17</sup> Ibid: 877.

<sup>18</sup> Ibid: 878.

then Rawls's remarks about the sense of justice would suffice to demonstrate the stability of the whole theory, and all the additional machinations of political liberalism would be unnecessary.

Barry contends that the congruence argument arises in the theory only because of an unfortunate eccentricity in Rawls's views about motivation. In his discussion of congruence, Rawls explicitly rejects what he calls the doctrine of the "purely conscientious act," associated with David Ross and H. A. Prichard. Barry argues that Rawls understands this doctrine to suggest that the recognition of the rightness of an action provides an overriding motive to perform it. Rawls worries that the claim makes such actions seem unintelligible or capricious: "on this interpretation the sense of right lacks any apparent reason; it resembles a preference for tea rather than coffee."<sup>19</sup> In contrast, Barry contends that in fact Prichard's claim amounts to "the thoroughly commonsensical idea represented by saying that people can do their duty out of a sense of duty and not in order to achieve some independently definable end."<sup>20</sup> At stake here is only the "quite natural" view, developed most thoroughly by T. M. Scanlon, that "the thought that something is the right thing to do is what motivates us to act rightly."<sup>21</sup>

From Barry's perspective, Rawls's rejection of the doctrine of the purely conscientious act is the first domino in a series that leads all the way to political liberalism. Barry suggests that because he rejects the idea that we can be motivated to act rightly just because we recognize that it is the right thing to do, Rawls must commit himself to an insanely high standard of moral motivation where the right action is always good for the agent. For if rightness alone is insufficient to motivate an action, then something else, an agent's own interests say, must be doing the motivating. But this, in Barry's view, leads to absurd conclusions. It forces Rawls to conclude that, for example, if I could not justly afford to take a vacation that I thought would be good for me, "I must somehow persuade myself that it would not be for my good at all. For only that thought can motivate me to refrain from taking the trip unjustly if the opportunity should arise."<sup>22</sup>

The only solution then is to appeal to the congruence argument and the Kantian interpretation, which attempts to show that the right action always promotes an agent's good. But by retaining the "thoroughly commonsensical idea . . . that people can do their duty out of a sense of duty and not in order to achieve . . . some independently definable end," Barry contends that the congruence argument would no longer be necessary and that consequently there is no need to appeal to such comprehensive Kantian elements. Justice as fairness is by itself both stable

---

<sup>19</sup> Rawls (1971): 478.

<sup>20</sup> Barry (1995): 884.

<sup>21</sup> *Ibid.*

<sup>22</sup> *Ibid.*: 889.

and political (in the technical sense), and there is no need to appeal to the family of ideas introduced in *Political Liberalism* in order to show that this is so.<sup>23</sup>

Barry's accusation, however, is surprising given Rawls's consistent commitment to the idea that duty itself is capable of motivating us to act. In *Theory*, Rawls frequently affirms the idea that the sense of justice can motivate us to action, claiming that while the suggestion may seem strange, "we . . . have the desire to act from a conception of right and justice."<sup>24</sup> In *Political Liberalism*, Rawls writes that "in setting out justice as fairness we rely on the kind of motivation that Scanlon takes as basic."<sup>25</sup> Barry takes Scanlon's explanation—that the desire to act in accordance with justice comes from a desire to act according to rules that could not reasonably be rejected by others who were similarly motivated—to be the best account of how we are in fact motivated by justice. This points to no disagreement with Rawls.

What then is Rawls denying in his critique of Ross and Prichard? Far from denying that we can act for the sake of duty, Rawls objects to their particular intuitionist claim that nothing can be said to explicate or define the nature of justice. This is what makes these actions look like inexplicable personal preferences. In this regard, both Rawls and Barry think that we are "moved by the aim of finding principles that others similarly motivated could not reasonably reject."<sup>26</sup> Indeed, as Samuel Freeman points out, there is something particularly curious about Barry's accusation, as the genealogy of this account of motivation runs not from Scanlon to Barry, but from *Rawls* to Scanlon to Barry.<sup>27</sup>

Where, then, does this leave us with regard to questions about the importance of Rawls's congruence argument to the question of political stability? Freeman suggests that something like the congruence argument is still necessary to resolve Rawls's worries about political stability. While institutions ordered according to justice as fairness might well encourage a sense of justice without the assurance of congruence, there is no guarantee that individuals can and should abide by the demands of justice when it appears to conflict with their own good. He nevertheless concludes that the original congruence argument fails to succeed on its own terms because it would not engender stability in a pluralistic society, and that the modifications of political liberalism do little to resolve the underlying problem. For, even if it is true that justice as fairness can serve as the object of an

<sup>23</sup> And so Barry concludes, "I believe that as time goes on, *A Theory of Justice* will stand out with increasing clarity as by far the most significant contribution to political philosophy produced in this century. Only one thing threatens to obscure that achievement: the publication of *Political Liberalism*," 915.

<sup>24</sup> Rawls (1971): 476.

<sup>25</sup> Rawls (1996): 50, n. 2.

<sup>26</sup> T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: the Belknap Press, 1998), p. 5.

<sup>27</sup> Freeman (2003): 310, n. 13.

overlapping consensus and that its internal dynamics of reciprocity encourages adherence to it by various reasonable comprehensive conceptions of the good, it is nevertheless the case that it apparently has nothing to say about why individuals ought to adhere to the demands of justice even when they conflict with their own good. In political liberalism, the motivational work that the congruence argument is supposed to perform gets exported to the particular reasonable comprehensive perspectives that people possess in a well-ordered society. But, as Freeman notes, “so far as overlapping consensus goes, justice may be no more than an instrumental good for many people and hence subject to compromise when it conflicts with their final ends.”<sup>28</sup> Reasonable comprehensive views *might* have their own arguments for congruence that will articulate reasons why justice ought to prevail in these situations, but there does not appear to be any reason to think that most or even many of these comprehensive views *will* provide this sort of assurance. These issues lead Freeman to conclude that the problem of stability to which congruence was meant to be a solution was exaggerated in *Theory*. He writes, “Rawls’ thought here may be that, since justice nonetheless occupies a significant position in each person’s view, even if only an instrumental good for many, whatever conflicts with their final ends there are will not be so frequent and entrenched as to undermine stability.”<sup>29</sup> Thus, Freeman admits that “the only way around this problem is to concede that the Kantian congruence argument is not needed to show the stability of a well-ordered society.”<sup>30</sup> In this way, he concedes that Barry must be right in his assertions that congruence is unnecessary, though for very different reasons.<sup>31</sup>

The problem of congruence therefore presents a double bind for political liberalism. If it is a genuine problem, it is one that is never resolved, even in Rawls’s later work. The original solution presented in *Theory* would actually undermine stability because it could be rejected by reasonable conceptions in a pluralistic society. The additional modifications of political liberalism provide no guarantee that peoples’ reasonable comprehensive views themselves contain a solution to the problem of congruence. Alternatively, if the problem of congruence is not a genuine problem, then all of the additional modifications of political liberalism are unnecessary because there was in fact never any stability problem in *Theory* to begin with. Political liberalism and all its innovations, including the liberal principle of legitimacy, appears to be either a failed solution to a real problem, or an unnecessary solution to an imaginary one.

---

<sup>28</sup> Ibid: 308.

<sup>29</sup> Ibid.

<sup>30</sup> Ibid: 306.

<sup>31</sup> In a footnote, Freeman adds, “So here we must partially concur with Barry’s criticism that the congruence argument is unnecessary but not for the same reasons he discusses, namely, that having a moral sense of justice is sufficient to prove stability. For reasons he discussed in the text, Rawls does not accept this; the problem of showing the rationality of justice still remains.” Ibid: 315, n. 58.



This understanding of political liberalism and the problem of stability, however, ignores the way in which Rawls's understanding of stability evolves in his later work. While it might be the case that Rawls retreats on the claim that the original problem of congruence poses a genuine threat to political stability, he comes to see a new problem, related to the issues of political stability that political liberalism aims to address. This is the problem of *legitimacy*.

## II. FROM STABILITY TO LEGITIMACY

One thing that can be said of Freeman's and Barry's readings of these sections of *Theory* is that they both take political stability to be a purely practical issue dealing with the feasibility of a conception of justice. Justice refers to the rights and obligations that exist between citizens and the state; stability is a related but separate practical problem. Within *Theory*, the content of justice is determined first, independent of any considerations regarding political stability. After the principles of justice are determined, the questions of political stability are raised as a separate second stage test to determine the feasibility or practicality of this conception of justice.<sup>32</sup>

Within Rawls's later work, however, this understanding of stability seems to evolve. While *Political Liberalism* is concerned with resolving these old issues, it is clear that the later work not only offers new solutions to old problems, but understands the problems themselves in new ways. *Political Liberalism* seems less concerned with the *practical problem of political stability* as it was originally understood, and more concerned with the *moral problem of legitimacy*.

In *Political Liberalism*, Rawls continues to argue that justice as fairness is best worked out as a two-stage process. He says that, "In the first stage it is worked out as a freestanding political (but of course moral) conception for the basic structure of society. Only with this done and its content—its principles of justice and ideals—provisionally on hand do we take up, in the second stage, the problem whether justice as fairness is sufficiently stable."<sup>33</sup> Rawls also continues to understand the question of stability to involve two primary issues. The first preserves his earlier discussion of the importance of a sense of justice, "whether people who grow up under just institutions (as the political conception defines them) acquire a normally sufficient sense of justice so that they generally comply with those

---

<sup>32</sup> As Rawls says in *A Theory of Justice*, "These remarks [on stability] are not intended as justifying reasons for the contract view. . . . At this point we are simply checking whether the conception already adopted is a *feasible* one and not so unstable that some other choice might be better," 504 (emphasis added).

<sup>33</sup> Rawls (1996): 140–1.

institutions.”<sup>34</sup> However, where *Theory* understands the second issue to involve the problem of congruence, Rawls now argues that it concerns the possibility that justice as fairness can serve as the focus of a shared overlapping consensus, “whether in view of general facts that characterize a democracy’s public political culture, and in particular the fact of reasonable pluralism, the political conception can be the focus of an overlapping consensus.”<sup>35</sup> Congruence is here conspicuously absent.<sup>36</sup>

If *Theory*’s worries about stability focus on whether people living under institutions ordered according to justice as fairness will at least regularly comply with its demands, the discussion in *Political Liberalism* is substantially different. In *Political Liberalism*, Rawls insists that the problem of stability, in sharp contrast to his earlier statements, is not meant to be understood as a “purely practical matter.” Rawls’s aim in demonstrating the stability of justice as fairness is *not* merely to show that “if a conception [of justice] fails to be stable, it is futile to try to realize it,”<sup>37</sup> or “to find ways to bring others who reject it also to share it; or failing that, to act in accordance with it, if need be prompted by penalties enforced by state power.”<sup>38</sup> Rather, what counts is “the kind of stability, the nature of the forces that secure it.”<sup>39</sup> More specifically, Rawls argues that the kind of stability that matters is intimately connected with its *legitimacy*. So he writes:

The point, then, is that the problem of stability is not that of bringing others who reject a conception to share it, or to act in accordance with it, by workable sanctions, if necessary, as if the task were to find ways to impose that conception once we are convinced it is sound. Rather, justice as fairness is not reasonable in the first place unless in a suitable way it can win its support by addressing each citizen’s reason, as explained within its own framework. Only so is it an account of the legitimacy of political authority as opposed to an account of how those who hold political power can satisfy themselves, and not citizens generally, that they are acting properly. A conception of political legitimacy aims for a public basis of justification and appeals to public reason, and hence to free and equal citizens viewed as reasonable and rational.<sup>40</sup>

This is just what Rawls calls the liberal principle of legitimacy, a substantially new preoccupation that begins to clarify some of the confusion in the earlier account of political stability. This is simply the idea that “since political power is the coercive power of free and equal citizens as a corporate body, this power should be

<sup>34</sup> Ibid: 141.

<sup>35</sup> Ibid: 140.

<sup>36</sup> It is a superficial, but perhaps telling fact that “legitimacy” does not appear in *A Theory of Justice*’s thorough index, just as “congruence” fails to appear in *Political Liberalism*’s.

<sup>37</sup> Rawls (1996): 142.

<sup>38</sup> Ibid.

<sup>39</sup> Ibid.

<sup>40</sup> Ibid: 143–4.

exercised, when constitutional essentials and basic questions of justice are at stake, only in ways that all citizens can reasonably be expected to endorse in the light of their common human reason."<sup>41</sup> From the perspective of *Political Liberalism* then, stability is not merely a concern with the practical feasibility of a conception of justice. Rather, it is connected with the deeper normative ideal that the coercive power of the state needs to be justified in some way to individual citizens.<sup>42</sup> Far from being merely a practical matter, it represents an important normative aspiration for any liberal theory of justice. As Rawls says, "If justice as fairness were not expressly designed to gain the reasoned support of citizens who affirm reasonable although conflicting comprehensive doctrines—the existence of such conflicting doctrines being a feature of the kind of public culture that liberal conception itself encourages—it would not be liberal."<sup>43</sup>

While *Theory* raises neither a general interest in legitimacy, nor its specific relation to the issue of political stability, the vantage and perspective that this new concern provides on the complicated question of political stability are illuminating, and help to explain a vital deficiency in Barry's attempt to free justice as fairness from the baggage of political liberalism. As we have seen, the failure of the congruence argument and the apparent impossibility of finding a suitable replacement for the work that it did seem to suggest that either the stability of a liberal regime is futile or else its stability can simply be assured through Rawls's arguments about the capacity of justice as fairness to encourage a sense of justice. Nevertheless, it seems clear that these new worries about liberal legitimacy introduce a set of considerations that Barry does not consider. Barry takes the problem of political stability to be a practical matter. To show that a conception of justice is stable is just to show that it is possible, that it is not some utopian hope to think that justice might be satisfied for people like us in this world. While this is perhaps the best way to understand the argument as it is set forth in *Theory*, it seems clear that the modifications offered in *Political Liberalism* not only attempt to address old, lingering problems, but reinterpret them in new ways. Thus, from the vantage of *Political Liberalism*, political stability is not only a practical problem or a futility test. Rather, it involves normative worries about the moral justification of coercive political power. Perhaps it is true that Rawls no longer thinks that demonstrating the congruence of justice and an individual's good is necessary to establish the feasibility of justice as fairness. Nevertheless, the new emphasis on the importance of the liberal principle of legitimacy raises new issues that cannot be solved by simply appealing to a conception of justice's capacity to inculcate a

---

<sup>41</sup> Ibid: 140.

<sup>42</sup> Consider also Rawls's remarks in *Justice as Fairness: A Restatement* (Cambridge, MA: The Belknap Press of Harvard UP, 2001), 184–6.

<sup>43</sup> Rawls (1996): 143.

sense of justice. For while the original theory may, in fact be sufficiently stable in the sense of being practical and feasible as a conception of justice, it does not wrestle in any way with the problem of political legitimacy.

It is important here to note that Rawls has not entirely changed the subject, raising a new issue and abandoning the initial problem of the futility or possibility of a conception of justice. For while legitimacy is a substantially new preoccupation, Rawls seems to think that the new concern with legitimacy and the additional features of political liberalism are the best way to satisfy the older worry about stability. For a legitimate political regime—one that wins “its support by addressing each citizen’s reasons, as explained within its own framework” and whose institutions are endorsed by an overlapping consensus of differing but reasonable views—is as stable as one could hope. It is a regime that reveals itself to be in accord with a citizen’s own deepest intuitions about politics and justice. The demand for legitimacy is moral; the result is practical.

### III. THE LIBERAL PRINCIPLE OF LEGITIMACY

It does not take much scrutiny to see how peculiar an account of legitimacy this is—at least by typical Kantian standards. Legitimacy is typically taken to refer to the right of states to coercively enforce their laws, and the correlative obligation of citizens to obey them.<sup>44</sup> Most liberal accounts of legitimacy are informed by the kind of neo-Kantianism that Rawls himself popularized. So, for example, Charles Beitz has argued that a regime is legitimate “if it *would be* consented to by rational persons subject to its rule.”<sup>45</sup> Here we see the familiar idea of the hypothetical contract that associates the legitimacy of a regime with its adherence to ideal principles of justice.

In contrast, Rawls’s new account strikingly abandons this purely hypothetical character. The legitimacy of a regime is not simply assured by its conformity to the appropriate principles of justice. What matters is not that “those who hold political power can satisfy themselves, and not citizens generally, that they are acting

---

<sup>44</sup> Though a growing number of recent works have challenged the idea that the right of states to enforce their laws ought to be understood to correlate with any obligations possessed by citizens. See for example, Allen Buchanan, “Political Legitimacy and Democracy,” *Ethics* 112 (July 2002): 689–719, 689–703; David Copp, “The Idea of a Legitimate State,” *Philosophy and Public Affairs* 28 (1999): 3–45, 10–16; William Edmundson, “Legitimate Authority without Political Obligation,” *Law and Philosophy* 17 (1998): 43–60, 44, 55–60; Robert Landenson, “In Defense of a Hobbesian Conception of Law,” *Authority*, ed. Joseph Raz (New York: New York UP, 1990), 36–37; and Christopher Heath Wellman, “Toward a Liberal Theory of Political Obligation,” *Ethics* 111 (July 2001): 735–59, 741.

<sup>45</sup> Charles Beitz, *Political Theory and International Relations* (Princeton: Princeton UP, 1999), 80.

properly.<sup>46</sup> Instead, legitimacy rests in the actual views of citizens and their ability to endorse the principles of justice that underlie a regime. A conception of justice must be able to “win its support by addressing each citizen’s reason, as explained within its own framework.”<sup>47</sup> What matters then is not merely the objectivity or soundness of the principles that organize a political regime (just though they may be), but citizens’ actual ability to endorse them.

This might seem like an unusual way to read Rawls on this point, as it seems to deemphasize the priority of the criterion of reasonableness and the two principles of justice. To be sure, not just any kind of endorsement of one’s political institutions matters. The views that have significance are limited to those that are reasonable. But it seems clear in these remarks that what matters with regard to legitimacy is not that the political institutions are *in fact* reasonable—that there is an objective normative justification for them. Rather, what matters is that the citizens themselves can be satisfied from the perspective of their own comprehensive views, and that they can themselves endorse them. It is hard to make sense of Rawls’s appeal to the importance of a society’s public political culture if the endorsement that matters is purely hypothetical. Public political culture matters because it plays such an important role in helping to shape the real views of citizens living in that culture.

For Rawls, the demand to justify the coercive power of the state to citizens in terms that they can accept is a matter of liberal principle. As he says “If justice as fairness were not expressly designed to gain the reasoned support of citizens . . . it would not be liberal.”<sup>48</sup> Rawls says little about the reasons that led him to reconsider the importance of legitimacy within an account of justice or the particular reasons that led him to abandon the kind of, hypothetical account that he himself popularized. It is easy enough, however, to see what is appealing about people’s actual endorsement of a political regime. For if the concern with legitimacy is born out of a respect for the freedom and equality of individuals and the need for the coercive power of the state to be justifiable to them, then it is not enough that the principles that underlie a regime can satisfy an objective but abstract standard of justice that citizens themselves do not accept.

This concern for the need to legitimate the coercive power of the state in personal terms is usually associated with voluntarism, a very different approach to these questions. John Simmons, for example, argues that the problem with the Kantian approaches that dominate much recent moral and political philosophy is that they are prone to confuse two separate categories of the moral evaluation of

---

<sup>46</sup> Rawls (1996): 143–4.

<sup>47</sup> *Ibid*: 143.

<sup>48</sup> *Ibid*.

political institutions.<sup>49</sup> On the one hand, we might talk about an institution's *justifiability*. This refers to certain *objective* claims that might be made to justify the institution in question. To justify the state would be to "show that one or more specific kinds of states are morally defensible."<sup>50</sup> A regime might be justifiable because it "is on balance morally permissible (or ideal) and that it is rationally preferable to all feasible nonstate alternatives."<sup>51</sup> A particular kind of state might be desirable because it meets certain objective standards of justice or attends to real human needs and welfare that could not otherwise be satisfied. The justification of a particular state or kind of state typically involves arguing, "that certain virtues that states possess or goods they may supply—such as justice or the rule of law—make it a good thing to have such states in the world."<sup>52</sup>

In contrast, legitimacy is a *subjective* or *agent-centered* concept that refers to the particular relationship between a citizen and her state. A state's legitimacy comes vis-à-vis its individual citizens, and therefore a state can be legitimate with respect to some but not all of the persons living under its rule. It refers to "the complex moral right [a state] possesses to be the exclusive imposer of binding duties on its subjects, to have its subjects comply with these duties, and to use coercion to enforce the duties. Accordingly, state legitimacy is the logical correlate of various obligations, including subjects' political obligations."<sup>53</sup> In short, a legitimate state has the right to rule its citizens.

For Simmons, then, justification and legitimacy pick out two separate dimensions of moral evaluation. It is possible, therefore, that a state might be justifiable, but illegitimate (with respect to some or most of its citizens); unjustifiable and illegitimate; unjustifiable and legitimate; and both justifiable and legitimate. There can be perfectly just states that do much good, but that nevertheless do not possess the special right to impose duties on their citizens. Likewise, there can be states that are corrupt or immoral, but which nevertheless are legitimate with respect to some or all of their citizens. The significance of legitimacy as a dimension of the moral evaluation of political institutions is that it draws attention to the particularizing, duty-creating circumstances that justification does not bring into view. It is not enough to simply say that states are doing good. States do not automatically have the right to command and coerce those living under their rule.

Simmons argues that much contemporary political theorizing, dominated as it is by Kantian influences, tend to confuse these separate categories because, in

---

<sup>49</sup> John Simmons, "Justification and Legitimacy," *Justification and Legitimacy* (New York: Cambridge UP, 2001), 122–57. Originally published in *Ethics* 109 (1999), 739–71.

<sup>50</sup> *Ibid.*: 125.

<sup>51</sup> *Ibid.*: 126.

<sup>52</sup> *Ibid.*

<sup>53</sup> *Ibid.*: 130.

their reliance on hypothetical justifications, they typically fail to recognize the significance of what people actually choose and do. As Simmons says, “Kantians think of institutional evaluation in terms of what ought to be chosen by people—that is, in terms of the moral quality of institutions, what makes those institutions good (virtuous, just, etc.)—not in terms of people’s actual choices.”<sup>54</sup> Actual choices are brushed aside to make way for the reasonable choices, not of actual persons, but their hypothetical, ghostly counterparts. Our actual choices and commitments, however, have an undeniable moral significance that simply is not captured by the traditional Kantian categories. Kantian theorizing has a generic character that makes the moral constraints that are imposed on persons living under a particular political regime seem, from their perspective, impersonal and alien. What would it matter to me, that some abstract and hypothetical version of myself might affirm a regime that I myself reject? In contrast, “How we have actually freely lived and chosen, confused and unwise and unreflective though we may have been, has undeniable moral significance; and our actual political histories and choices thus seem deeply relevant to the evaluation of those political institutions under which we live.”<sup>55</sup>

As we have seen, the later Rawls is rather less characteristically Kantian than he is often taken to be, and the theory of legitimacy that arises in his later work offers an interesting, if puzzling, middle ground between Simmons’s voluntarism and Kantian hypothetical justifications. For it is clear that from the perspective of Rawls’s later work, purely hypothetical or abstract standards of justice are not sufficient to explain the legitimacy of a regime. It is not enough that political authorities can satisfy themselves that their coercive actions fulfill the appropriate standards of justifiability by pointing to objective principles of justice. Rather, a legitimate regime must be able to “win its support by addressing each citizen’s reason, as explained within its own framework.” What a person *actually* thinks matters a great deal. By rooting an account of political justice in the concepts latent within a public political culture, Rawls aims to generate principles of justice and political institutions that are rooted in ideals that are already accepted by the persons who will be subject to them.

Simmons himself recognizes that Rawls’s “political” turn attempts to offer “a more personal and uncontroversial justification for the use of political power.”<sup>56</sup> Simmons nevertheless argues that this Rawlsian middle ground is both “watered-down and one-sided.” The account is watered down because of its search “for a single compromise dimension of moral evaluation located somewhere between

---

<sup>54</sup> Ibid: 147.

<sup>55</sup> Ibid: 148–9.

<sup>56</sup> Ibid: 147.

impersonal justifications and personal legitimations,”<sup>57</sup> and thus becomes one sided in its inability to preserve a sharp distinction between justification and legitimacy.

Taking the second of these worries first, given the preceding interpretation of the aspirations of Rawls’s later work, I think we can say that the account is clearly not “one-sided.” As we have seen, Rawls’s insistence that political institutions ought to be justifiable to the persons living under them in terms that they can actually accept is a long way from a hypothetical justification. Unlike these accounts, there is an insistence that the satisfaction of objective and abstract standards is not sufficient, but that legitimacy must have a point of contact with actual peoples’ normative commitments. Thus, there is no automatic inference from justification to legitimacy. Rawls can, on the one hand, speak of certain objective criteria for evaluating political regimes—namely the two principles of justice—and on the other hand, inquire as to whether these principles can be justified to citizens in terms that they can accept. As we have seen, it is part of the structure of Rawls’s theory that inquiries into political legitimacy are external to questions of justice and operate as a second stage test of a conception of justice after its content has already been established. Justice and stability are clearly different qualities a regime might possess. In this way, he can preserve a sharp distinction between these two dimensions of institutional evaluation. It is one thing to say that a regime is just; it is quite another to say that it is legitimate.

Moving to Simmons’s first worry, what kind of legitimacy does this actually amount to and to what extent is it “watered down?” I think it is best to understand Rawls as being interested in a conception of legitimacy that Simmons would call “attitudinal.” On this approach, legitimate regimes are those “that are accepted or approved of by their subjects in certain distinctive ways.”<sup>58</sup> This is what Charles Taylor has in mind when he writes that legitimacy “designates the beliefs and attitudes that members have toward the society they make up. The society has legitimacy when members so understand and value it that they are willing to assume the disciplines and burdens which membership entails. Legitimacy declines when this willingness flags or fails.”<sup>59</sup> Legitimacy in this sense is

<sup>57</sup> Ibid: 148.

<sup>58</sup> Ibid: 131.

<sup>59</sup> Charles Taylor, “Alternative Futures: Legitimacy, Identity, and Alienation in Late Twentieth Century Canada,” *Communitarianism: A New Public Ethics*, ed. M. Daly (Belmont, CA: Wadsworth, 1994), 58. Quoted from Simmons (2001): 132, n. 24. Taylor distinguishes this attitudinal sense of legitimacy from one used as an objective evaluation of regimes. It is important to note that while Rawls’s interest in legitimacy is similarly attitudinal, he does not seem to be specifically interested in a “legitimation crisis” in Taylor’s sense.



an attitudinal concept referring to “members’ positive beliefs, attitudes, perceptions, or other “favorable orientations” toward their society or its regime.”<sup>60</sup> Such attitudes are quite distinct from consent since they do not involve any explicit contact scenario.<sup>61</sup> An attitudinal approach, however, is similar to consent accounts since these attitudes are actual and personal. Regimes are legitimated not by their abstract, hypothetical properties (important though these may be), but because actual people really endorse them (even if they do not consent to them).

Simmons argues that the general problem with such attitudinal accounts is that they are problematically subjective. They “turn out to be about the wrong thing,” making “judgments of legitimacy too much about subjects and too little about their states. To call a state legitimate is surely to say something about *it*, about the rights it possesses or the scope of its authority.”<sup>62</sup> In focusing on the subjective attitudes of citizens, attitudinal accounts leave it open that “states could create or enhance their own legitimacy by indoctrination or mind control; or states might be legitimated solely by virtue of extraordinary stupidity, immorality, imprudence, or misperceptions of their subjects. Surely none of this is what any of us has in mind when we call a state or a government legitimate.”<sup>63</sup>

Attitudinal accounts, however, seem no more problematically subjective than consent-based accounts of legitimacy. For voluntarists like Simmons, legitimacy refers to a property of states (namely, their right to impose obligations on their subjects) that they possess in virtue of something about their subjects (namely, that they consent to their rule). Similarly, attitudinal accounts focus on a property of states that they possess in virtue of something about their citizens (namely, that they endorse their political institutions). Simmons’s worry is that attitudinal accounts lose track of the dubious ways in which states might gain their citizens’ endorsement. In this regard, however, there appears to be no difference between consent-based approaches and attitudinal ones. For voluntarists, consent is only morally binding when it is given under appropriate conditions. If I grant my consent to a regime because I am mad or incompetent, or have been duped, bribed, coerced, or otherwise cowed into submission, then such a regime has no legitimate authority over me. Similarly, a person’s endorsement

---

<sup>60</sup> Simmons (2001): 133.

<sup>61</sup> Of course, however attractive consent might be as a normative ideal, it is unworkable in practice since virtually no one consents to be ruled by their states under the kinds of conditions that would make such consent normatively significant. For more on this, see A. John Simmons, *Moral Principles and Political Obligations* (Princeton: Princeton UP, 1979), ch. 3 and 4.

<sup>62</sup> *Ibid.*

<sup>63</sup> *Ibid.*: 134.

or affirmation of a regime would be morally insignificant if it was achieved by trickery, deceit, coercion, inducements, manipulation, or other dubious methods. Attitudinal accounts are no more problematically subjective than consent theories of legitimacy.<sup>64</sup>

While an attitudinal account of legitimacy does not suffer from the internal incoherence that Simmons alleges, it will still seem watered down from the perspective of a voluntarist account of legitimacy and political obligation. The appeal of voluntarist accounts of political legitimacy comes from the fact that they take individual liberty so seriously. The coercive power of the state stands in need of justification, and my liberty is so weighty that it can only be ceded through an act of consent. Consent makes our political commitments “seem less external and more obvious . . . [and] makes the constraint more likely to be motivationally efficacious.”<sup>65</sup> Objective justifications of state power that do not directly involve those subject to it are, in this way, beside the point. The moral considerations that make legitimacy matter demand that it can be achieved only through that which is most “direct and personal.”

Rawls’s liberal principle of legitimacy is rooted in the same normative concerns. Like voluntarist accounts, it begins with what ought to be a properly liberal worry about the need to explain the appropriateness of a state’s coercive political power to those subject to it. Like voluntarist accounts, it suggests that any justification of coercive political power needs to be actual and personal, rather than abstract and hypothetical. The endorsement of one’s political institutions on the basis of a reasonable comprehensive view is, after all, an actual, direct, and personal subjective state. Because of this, citizens who endorse their regime will be motivated to support it and obey its laws, for the institutions of such a regime reflect the political ideals of the citizens under its rule. Their deepest-held political ideals are instantiated in its political structures. Endorsed regimes, even as they exercise coercive political power, will not appear as a foreign or alien imposition into the lives of their citizens, but rather will represent the common political ideals that they hold most dear. Thus, the citizens of an endorsed state are already deeply committed to the norms behind their political institutions. They are precisely the institutions that they want, and they are therefore motivated to take on the burdens and responsibilities of membership.

---

<sup>64</sup> This, of course, leaves it open that rational, fully competent people might endorse authoritarian regimes. This might seem worrisome, but it is important to remember that while it might grant such a regime legitimacy with respect to that person, that does not mean that such a regime is just or otherwise immune from criticism. And the authority it wields is binding only insofar as an individual continues to endorse it.

<sup>65</sup> Ibid: 147.

## IV. LEGITIMACY AND RECONCILIATION

In this way, Rawls's liberal principle of legitimacy offers a highly novel way of thinking about coercive state authority, and reveals his late interest in Hegel's notion of reconciliation as the task of political philosophy.<sup>66</sup> Rawls is not merely interested in how state power is licensed by individuals or in narrow questions about rights and obligations. He is also not interested in simply illuminating the generic benefits that states provide. On Rawls's account, citizens of a state that has earned their endorsement have not just licensed it to rule, but they are *reconciled* to it, and embrace its authority since it is structured according to the moral ideals to which they are already committed.

For Hegel, human freedom is only fully realized when our social and political institutions are structured in such a way as to enable human freedom, and this fact is recognized by human beings. Thus, human freedom is not fully actualized until we are self-conscious about it, and understand how and why our political institutions enable our own freedom. The task of political philosophy is not merely to engage in the academic exercise of justifying an account of social justice. Rather, it is practical. As Rawls says of Hegel, "When in our reflections we understand our social world as expressing human freedom and enabling us to achieve it as we lead our daily life, we become reconciled to it."<sup>67</sup> The task of political philosophy is to help us to see the way in which our political institutions are reasonable or rational, and when this is clear to us, we need not be simply resigned to them, but can embrace them as meeting our fundamental human needs. Thus, "it is only within a rational (reasonable) social world, one that by the structure of its institutions guarantees our freedom, that we can lead lives that are fully rational and good. And although no social world can guarantee happiness, it is only within such a world that full happiness can be attained."<sup>68</sup>

Within his own work, such talk is captured by the idea of a "realistic utopia." Rawls says that, "political philosophy is realistically utopian when it extends what are ordinarily thought of as the limits of practical political philosophy."<sup>69</sup> Such an approach aims at describing an ideal vision of a reasonable and rational liberal political society (or international society) that is nevertheless possible even if it is not actual. In doing this, such a philosophical project "[banishes] the dangers of resignation and cynicism . . . and in working toward it gives meaning to what we

---

<sup>66</sup> See, for example, *The Law of Peoples* (Cambridge: Harvard UP, 2001), 124–8; *Justice as Fairness: A Restatement* (Cambridge: Belknap, 2001), 3–4; and *Lectures on the History of Moral Philosophy* (Cambridge: Harvard UP, 2000), 331–6.

<sup>67</sup> Rawls (2000): 332.

<sup>68</sup> *Ibid.*

<sup>69</sup> Rawls, *The Law of Peoples* (2001): 6.

can do today.<sup>70</sup> When we are reconciled to our political institutions, we come to understand how they are reasonable and rational, and are able to embrace them.

In this way, political philosophy for Rawls is not merely a theoretical activity, but plays an important practical role in enabling a certain kind of political life. Citizens who are able to endorse their regimes on the basis of their own reasonable comprehensive philosophical commitments are reconciled to those regimes and see them as an expression of their own deepest normative commitments. One might say that under such an approach, the problem of legitimacy is not so much solved as dissolved. For those citizens who endorse their states, there is no question of whether the state has the right to coercively enforce the law and whether they are bound by correlative obligations to obey. Rather, they actively support the state, respect its institutions, and willingly do their part. Obligation is replaced by enthusiasm. Indeed, in this way endorsement might be seen as creating an even more direct and personal relationship between a citizen and her state than consent. For while genuine consent can create binding moral obligations, the mere presence of consent at some time in the past says little about peoples' actual current attitudes of support or enthusiasm for their state. Citizens may have consented to accept the authority of a political regime and thereby be bound by whatever duties were included in the terms of consent, but may come to see these duties as onerous, even if their state has not violated the terms of the agreement. They are still morally obligated, but their prior commitments will hardly seem less external, more obvious, and more likely to be motivationally efficacious. The deal will still be binding, but they will view the obligations of membership and the state's demands as alien, oppressive, and constricting. In contrast, citizens who are reconciled to their institutions do not merely see them as a necessary evil, or as the best alternative in a non-ideal world. Instead they see them as reasonable, as embodying their own normative commitments, and as making possible a world where their own happiness can be attained.

Earlier in this paper I suggested that one of the reasons why Rawls's liberal principle of legitimacy has failed to receive the attention that it deserves is that his critics have failed to properly appreciate the real normative point behind his shift to a political liberalism. This is perhaps understandable as Rawls is hardly as clear as one might hope in explicating the precise issues that led to his modifications. Indeed, I have argued that if the problem of political stability is understood as it is in *Theory* as a practical worry, then there is no stability problem to be resolved. The original account of justice as fairness is a feasible conception of justice. In *Political Liberalism*, however, the problem of stability becomes reinterpreted as an interest in legitimacy, understood as a kind of endorsement of political

---

<sup>70</sup> Ibid: 128.

institutions on the basis of citizens' own reasonable comprehensive moral views. The citizens of an endorsed regime do not view its laws and demands as an oppressive, alien imposition into their lives, but rather they are reconciled to these political institutions, seeing them as an instantiation of their own deepest moral commitments and happily undertaking the burdens of membership. In this way, Rawls's liberal principle of legitimacy offers an extremely novel approach to the problem of legitimating the coercive power of the state. A discussion of it also helps to illuminate the point and significance of his turn to a political liberalism.

It is often said that Rawls's turn to a political liberalism is motivated by a weakening of his own commitment to an objective liberal political morality, and an embrace of "communitarian" or "pluralist" values. The liberal principle of legitimacy, however, is rooted in liberal concern. The problem is not that Rawls has abandoned his commitment to liberalism, but rather that many liberal political theorists think that an objective, impartial justification of political authority suffices to explain the legitimacy of a regime. But for Rawls, something more is required. To put it crudely, it is a good thing that citizens can endorse their political institutions, seeing them not as an alien or foreign imposition into their lives, but as an instantiation of their own political ideals. It is a good thing because such citizens are reconciled to their political institutions and are happy to support them. Such institutions are as stable as one can hope for. Rawls's liberal principle of legitimacy is a good thing for what ought to be liberal reasons.<sup>71</sup>

*Louisiana State University*

---

<sup>71</sup> I am grateful for conversations and comments from Andy Altman, Richard Arneson, Tal Brewer, Ian Cruise, Samuel Freeman, Lucan Gregory, John Simmons, and Kit Wellman, as well as the comments from an anonymous reviewer for *The Philosophical Forum*.