

*Legitimacy and the Project of Rawls's Political Liberalism*<sup>1</sup>

Paul Weithman  
Department of Philosophy  
University of Notre Dame

In “Realism and Moralism in Political Theory,” Bernard Williams wrote that the John Rawls of *Political Liberalism*

wants to make a bigger gap than *TJ* allowed between two different conceptions: that of a society in which power is rightfully exercised (a well-ordered society), and that of a society that meets liberals’ aspirations to social justice.<sup>2</sup>

If we take the phrase “a society in which power is *rightfully* exercised” to mean “a society in which power is *legitimately* exercised”, then Williams’s remark suggests a reading of *PL* that is influential and widely endorsed. Indeed, I believe that in some form, this reading is often taken for granted. I shall refer to the reading I have in mind as “the standard reading”.

The standard reading proceeds in four steps.

- First, it purports to explain Rawls’s political turn. According to the standard reading, after publication of *TJ*, Rawls came to see that its arguments for the principles of justice rested upon Kantian claims that some members of the well-

---

<sup>1</sup> I am grateful to Dan Brudney, Ryan Hammond, Ben Laurence, Malcolm McCollum, Caleb Pearl, John Simmons and audiences at the University of Tennessee, Knoxville and the University of Chicago for helpful comments on earlier drafts.

<sup>2</sup> Bernard Williams, “Realism and Moralism in Political Theory”, in his *In the Beginning was the Deed* (Princeton University Press, 2005), pp. 1-17, p. 1. By ‘*TJ*’ Williams is clearly referring to John Rawls, *A Theory of Justice* (Harvard University Press, 1999). Like Williams, I shall refer to *Theory of Justice* by its initials and shall give page references parenthetically in the text; I shall refer to Rawls, *Political Liberalism* (Columbia University Press, 1995) as ‘*PL*’ and give parenthetical citations to it as well.

ordered society (hereafter, “the WOS”) could reasonably reject. Reasonably rejecting the arguments for the principles, they would reasonably reject the principles themselves. Rawls therefore recast justice as fairness as a political conception of justice, thereby weakening the premises on which his defense of the principles rested, so that the principles and their grounds would be acceptable to all reasonable members of a WOS.

- Second, Rawls came to see that even after justice as fairness had been recast, it – and hence the two principles of justice -- would not enjoy unanimous consent in the WOS. Instead, reasonable citizens of that society would adhere to a variety of liberal political conceptions of justice.
- Third, Rawls continued to think that the demands of a WOS must be publicly justifiable by a principle or principles acceptable to all reasonable citizens. In *TJ*, the two principles of justice were said to provide a “common standpoint” for imposing demands and adjudicating conflicting claims. (*TJ*, p. 4) Once Rawls realized that some citizens of the WOS would reasonably reject his two principles, he saw that he needed a weaker standard of public justification than he had defended in *TJ*. In *PL*, the liberal principle of legitimacy -- rather than the two principles of justice – provides that standard.

- Finally, because the principle of legitimacy replaces the principles of justice as the common standard of public justification, legitimacy rather than justice is the subject matter of *PL*.

As we shall see in the next section, the standard reading has much to be said for it.

Moreover, by showing why Rawls opens the gap to which Williams refers, it helps to explain the retreat from egalitarianism that some readers think they have detected in *PL*.

Despite the appeal of the standard reading, its central claims are as disturbing as they are striking. According to that reading, the Rawls of *PL* does not just correct and reformulate justice as fairness. He reorients his thought around a political concept that had no explicit place in his earlier work. In doing so, he is said to take back one of his boldest and most attractive moral commitments. If a reading with these implications is to be accepted, it ought to be accepted only if very strong textual and philosophical grounds can be found in its favor, and only after a thorough evaluation of interpretations that avoid them.

I shall argue that the standard reading errs in a number of places. It goes wrong at the first step by misidentifying the problem that led Rawls to recast justice as fairness. In moving from the first step via the second to the third, it overlooks an important fact: while the Rawls of *PL* *did* concede that members of the well-ordered society would endorse different conceptions of justice, he conceded it long after he began to recast justice as fairness. This fact is easy to overlook if we fail to distinguish as sharply as Rawls would between pluralism about the good, alleged at the first step,

and disagreement about justice, alleged at the second. But if we overlook it, we will mistake the central project of *PL* for a problem which attracted Rawls's notice quite late in the execution of that project: a problem of showing stability in the face of likely differences about the right. A consequence of this mistake will be a misleading description of *PL* – such as that offered at the fourth step of the standard reading.

On my reading, Rawls recast justice as fairness because he thought *TJ* failed to show that justice as fairness would be stable in a society characterized by reasonable pluralism about the *good*.<sup>3</sup> To the extent that *PL* has a single purpose, it is to correct that failure. The standard reading is right in claiming that Rawls introduced the principle of legitimacy to help correct it. As we shall see, the standard reading claims that when citizens recognize laws as legitimate, their recognition engages their sense of duty. That, it says, is how legitimacy contributes to stability and solves the problem with *TJ*. I shall argue that while this is *one* of the ways legitimacy contributes to stability, we will miss *PL*'s central project if we think it is the *only* way. For *PL*'s discussion of legitimacy does not add anything significant to what *TJ* had said about the duty of citizens in the WOS to obey the law. It recalls and labels what Rawls said about that duty in his earlier work, and applies it to cases which are not unlike those he

---

<sup>3</sup> It is often said that Rawls tried to show that a well-ordered society would be stable for the right reasons. This is misleading insofar as it suggests that Rawls was concerned to show that the persistence of basic of institutions. It would be more accurate to say that Rawls wanted to show that a well-ordered society would be “stably just”, as Joshua Cohen rightly points out at “Democratic Equality”, *Ethics* 99 (1989): 727-51, pp. 743-44.

The well-ordered society of *TJ* will be stably just only when terms of cooperation adopted in the original position are stably adhered to – that is, when its conception of justice is stable. When speaking in my own voice, I shall therefore refer to the stability of justice as fairness rather than the stability of a well-ordered society.

considered in *TJ*. But if *PL*'s treatment of legitimacy does not add to *TJ*'s discussion of this important element of the *right*, it does allow Rawls to offer a fresh and important argument about citizens' *good*. For showing that justice as fairness would be stable involves showing that citizens would regard doing their duty as part of their good. Rawls's treatment of legitimacy in *PL* helps him establish that conclusion on the basis of more realistic assumptions than he had used in *TJ*.

I shall defend this interpretation beginning in §4. I believe that it is important to see the project of *PL* aright because I think that Rawls has much to teach us about stability and legitimacy. But I shall not draw out the implications of his account of legitimacy here. My aim is simply to recover the account so as to gain more accurate views of legitimacy's role in *PL* and of *PL*'s central project. As we shall see, this requires a surprising amount of rational reconstruction, since much of the work legitimacy does is compressed into Rawls's discussion of a couple of difficult cases.

Before I defend my own interpretation, I shall elaborate the standard reading. That reading is not, to my knowledge, explicitly laid out and defended in the literature on *PL*. But many readers of *PL* have offered criticisms and interpretations that can plausibly be seen to stem from a single underlying reading of the book. In §1, I surface that interpretation and make good my claim about the pervasiveness of its hold by showing how it unites and explains some important commentary on *PL*. Because of the influence of this reading, it merits careful attention. In §§2 and 3, I defend my claim that the reading goes wrong at each of the four steps laid out above. As we shall see at the end of §3, showing the problems with the standard reading does not just clear the

way for a more faithful interpretation of *PL*. It also brings to light a number of philosophical and textual puzzles about legitimacy that a correct reading must solve. In §§6 and 7, I show how the reading defended here addresses those puzzles.

### *§1. The Standard Reading*

I observed at the outset that the standard reading is suggested by a remark of Bernard Williams's. Elements of the reading are laid out more expansively by other writers.

I said that the first step of the standard reading is to explain Rawls's political turn by pointing to the possibility that some members of the well-ordered society would reasonably reject Rawls's two principles. Consider Burton Dreben's description of the "essential flaw" Rawls saw in *TJ* and the reason for his shift to political liberalism:

the last third of the book *A Theory of Justice* deals with th[e] question of stability, or as Rawls comes to call it in much later writings "stability for the right reasons." And the way he argues that the two principles of justice, which the first two-thirds of the book deals with and which are to govern the basic structure of society, are indeed stable . . . rests on showing that everyone will agree, or at least the vast majority of the society will agree, on these principles of justice. Now what Rawls began to see was that, under the very conditions that satisfy the principles of justice that he worked so hard to establish, reasonable and free and equal people will begin to differ, inevitably and properly so, on those very principles of justice. Hence, from his perspective, the theory of stability that he had set forth in the last third of the book contradicts the first two-thirds of the book. This leads to a recasting of what he became so world famous for."<sup>4</sup>

This passage reaches a striking conclusion: Rawls recast justice as fairness because the first two-thirds of *TJ* contradict the last third. Striking as it is, this

---

<sup>4</sup> Burton Dreben, "On Rawls and Political Liberalism", *The Cambridge Companion to Rawls* (Cambridge University Press, 2003) ed. Freeman, pp. 316-46, pp. 316.

conclusion echoes Rawls's own claim that he took his political turn because "the account of stability in part III of *Theory* is not consistent with the view as a whole." (*PL*, pp. xvii-xviii) For present purposes, what is most interesting about the passage is what Dreben says about *how* the first two-thirds of *TJ* contradict part III. He says the contradiction stems from the fact that members of the WOS would differ about the principles of justice and that their differences would be "inevitabl[e] and proper[.]". Thus Dreben seems to think that some members would reasonably reject the principles and that the contradiction which stems from their doing so is what "leads to a recasting" of justice as fairness.

Dreben does not say *why* members of a well-ordered society would "differ on [Rawls's] principles". It is natural to suppose that he thinks they would differ because some of them would reject Rawls's arguments for the principles, including the argument from the original position that Rawls offers in "the first two-thirds of the book". The question, then, is why Dreben thinks they would reject those arguments. I said earlier that according to the standard reading, some of members of the WOS would reject them because those arguments rest on Kantian premises that some reasonable citizens would find objectionable. Dreben hints at that explanation<sup>5</sup> and Bruce Ackerman – whose writings express elements of the standard reading -- offers an extended argument to that effect.<sup>6</sup> I shall take Ackerman's argument as an elaboration of the hint dropped by Dreben, and hence as an explanation of the disagreement to

---

<sup>5</sup> Dreben, "Rawls and Political Liberalism", p. 345.

<sup>6</sup> Bruce Ackerman, "Political Liberalisms", *The Journal of Philosophy* 91, 7 (1994): 364-86, pp. 371-75.

which Dreben calls attention. If this is a fair reading of Dreben, then he quite clearly takes the first step of the standard reading.

According to the second step, Rawls came to see that members of the WOS would disagree about justice as fairness even after it was recast as a political liberalism. About this step, Dreben is considerably less clear. Immediately before the passage quoted just above, he writes

The first book deals with justice, a much discussed topic; the second book deals with legitimacy, a topic that few contemporary philosophers in the liberal tradition have focused on. (It of course has been dealt with by various so-called political scientists.) The question of legitimacy – that is, under what conditions will someone properly accept a law as legitimate, even if he differs with it, even if he thinks it unjust – is a central question for present-day society. And that is what Rawls is really considering. It grows out of what he considers to be an essential flaw in the first book[.]<sup>7</sup>

Dreben's conjunction of the two quoted passages suggests that he moves directly from the first step of the standard reading to the third, which asserts that because of disagreements about justice in a well-ordered society, legitimacy replaces justice as the standard of public justification. But clearly the first and second steps need to be distinguished since, once justice as fairness has been recast as a political liberalism, Rawls's principles no longer depend upon Kantian premises. The reasons for the rejection of the principles asserted at the first step no longer obtain and Dreben's move to the third step is too quick.

The second step is more explicitly taken by other commentators. Thus when David Reidy says that in a well-ordered society, citizens would adhere to "any number

---

<sup>7</sup> Dreben, "Rawls and Political Liberalism", p. 316.

of possible (generically liberal) reasonable conceptions of justice”, context suggests that he means they would adhere to any number of liberal *political* conceptions of justice.<sup>8</sup> When Simon Cabulea May says that the Rawls of *PL* “assumes the possibility of widespread disagreement about justice”,<sup>9</sup> what he says later makes clear that he, too, is talking about disagreement over liberal political conceptions of justice.<sup>10</sup> And when David Estlund says “[i]t is impossible to deny that Rawls holds in *PL* that there are many political liberalisms, justice as fairness being but one” he quite obviously means that Rawls thinks there are many political liberalisms among which citizens of a WOS would reasonably divide.<sup>11</sup>

But if Reidy, May and Estlund all separate more sharply than Dreben does the claims that I have identified as the first and second steps of the standard reading, Reidy and May say things which make Dreben’s elision of the two steps understandable. Reidy says “freedom leads to reasonable disagreement over matters of religious, philosophical and moral doctrine, as well as over theories of justice”.<sup>12</sup> As we saw, reasonable disagreement about comprehensive doctrines is what led to the possibility, asserted at the first step, that some citizens of the well-ordered society of *TJ* would

---

<sup>8</sup> David A. Reidy, “Reciprocity and Reasonable Disagreement: From Liberal to Democratic Legitimacy”, *Philosophical Studies* 132 (2007):243-291, p. 251.

<sup>9</sup> Simon Cabulea May, “Religious Democracy and the Liberal Principle of Legitimacy,” *Philosophy and Public Affairs* 37,2 (2009): 136-70, p. 148.

<sup>10</sup> May, “Religious Democracy”, p. 155.

<sup>11</sup> David Estlund, “The Survival of Egalitarian Justice in John Rawls’s *Political Liberalism*”, *Journal of Political Philosophy* 4 (1996): 68-78, typescript pp. 4, 12.

<sup>12</sup> Reidy, “Reciprocity and Reasonable Disagreement”, p. 247.

disagree with Rawls's two principles as defended in that book. What I take Reidy to be suggesting is that the same dynamics that open that possibility would open the possibility, asserted at the second step, that citizens of the well-ordered society of *PL* would dissent from justice as fairness when it is presented as a political liberalism. May seems to read Reidy this way and to accept his suggestion.<sup>13</sup> If the disagreements about justice asserted at the first and second steps have a common cause, as Reidy and May seem to think, then it is more understandable that Dreben simply asserted that Rawls realized there would be disagreements about justice in the well-ordered society and moved from that assertion to the third step. I shall argue shortly that the assertion of a common cause of the two instances of disagreement is, at best, highly misleading.

I have already noted that Dreben takes the third step of the standard reading. Reidy suggests a move from the second to third step when he says "No longer able reasonably to assume the sort of general consensus he thought likely to be ushered in by the institutional embodiment of his two principles, Rawls found issues of legitimacy increasingly central to his project."<sup>14</sup> In Estlund, the move from the second step to the third is especially clear.<sup>15</sup>

Dreben, Reidy, May and Estlund all move from the third step of the standard reading to the fourth, asserting that as justice forms the subject matter of *TJ*, so legitimacy forms the subject matter of *PL*. Thus as we saw in the second passage I

---

<sup>13</sup> May, "Religious Democracy", p. 155.

<sup>14</sup> Reidy, "Reciprocity and Reasonable Disagreement", p. 247.

<sup>15</sup> Estlund, "Egalitarian Justice", typescript pp. 6ff.

quoted from his essay, Dreben says that Rawls's "first book deals with justice, a much discussed topic; the second book deals with legitimacy".<sup>16</sup> Reidy says that legitimacy is "center stage" in *PL*, "pushed" there by the pluralism that he takes to explain the disagreement asserted at the second step and to necessitate the introduction of legitimacy at the third.<sup>17</sup> May says that in moving from *TJ* to *PL* Rawls "shift[s] from the problem of distributive justice to the problem of political legitimacy".<sup>18</sup> Estlund says that legitimacy is what *PL* is "primarily about".<sup>19</sup>

Readers who take what I have identified as third step of the standard reading differ about the reasons for taking it. These differences give rise to two variants of the standard reading. On one variant, which I shall call the "foundationalist variant",<sup>20</sup> a commitment to the public justifiability of basic political arrangements formed the foundation of Rawls's thought, both early and late.<sup>21</sup> In *TJ*, he argued that a society could honor that commitment only if it conformed to the two principles of justice. Once he realized that the principles were too strong, he replaced them with the weaker principle of legitimacy, again to show what standard a society must meet if it is to honor that foundational commitment.

---

<sup>16</sup> Dreben, "Rawls and Political Liberalism", p. 316.

<sup>17</sup> Reidy, "Reciprocity and Reasonable Disagreement", p. 247.

<sup>18</sup> May, "Religious Democracy", p. 148.

<sup>19</sup> Estlund, "Egalitarian Justice", p. 68.

<sup>20</sup> I am grateful to John Simmons for urging me to discuss this variant.

<sup>21</sup> Charles Larmore, *The Autonomy of Morality* (Cambridge University Press, 2008), pp. 147ff.

This foundationalist variant has considerable appeal, for Rawls *does* seem to be fundamentally committed to public justifiability. Indeed, that commitment may seem to be what makes Rawls's view paradigmatically liberal.<sup>22</sup> Of course, fully to spell out this variant, we would have to know how the commitment to public justifiability is to be understood. The natural way to take the commitment is deontically, as expressing or stemming from a requirement. This is, I believe, the way Ronald Dworkin understood the commitment in an early and very influential essay on *TJ*.<sup>23</sup> If Rawls thinks the commitment is a requirement of political morality, then he must also think that the requirement can be represented as the object of a choice in the original position. But because the commitment says that principles governing the exercise of political power must be justifiable to persons as free and equal, the commitment seems to *underlie* the requirement that basic principles of political morality be chosen in the original position, since the original position is used to identify principles that are justifiable to persons as free and equal. The foundationalist variant of the standard reading therefore seems to require that we read Rawls as arguing in a circle. And that is precisely how Charles Larmore interprets Rawls, in what I read as the clearest and most forceful defense of the variant.<sup>24</sup>

---

<sup>22</sup>Jeremy Waldron, "Theoretical Foundations of Liberalism," *The Philosophical Quarterly* 37 (1987):127-150.

<sup>23</sup> Ronald Dworkin, "The Original Position," in *Reading Rawls*, ed. Norman Daniels. Oxford: Basil Blackwell, 1975, pp. 16-52.

<sup>24</sup> Larmore, *Autonomy of Morality*, pp. 151ff.

To rebut the foundationalist variant, it is necessary to spell out Rawls's claim that justice as fairness is based on a set of "conceptions" or "ideals" rather than on a basic right or on a requirement of mutual justifiability.<sup>25</sup> I have tried to do that elsewhere and shall not rehearse my interpretation here.<sup>26</sup> For now, suffice it to say that since the foundationalist variant involves imputing a circularity to Rawls, it should be accepted only if no other reading of the texts is available. Because I believe another and a more defensible reading *is* available, I shall confine my attention to the other variant of the standard reading, which takes the third step for different reasons. That variant is suggested by Dreben.

According to Dreben's variant, Rawls thought that citizens of a well-ordered society would feel duty-bound to honor their society's demands only if they could see that those demands were justifiable by a common standard. Since a widespread and well-founded sense of duty is necessary for the right kind of stability, Rawls needed the principle of legitimacy to provide a common standard, elicit citizens' sense of duty and show how the well-ordered society would be stabilized "for the right reasons". If it is fair to impute the standard reading to Dreben at all, it is this variant that we have to impute to him. For the two extracts from his essay – taken together -- make clear that he thinks the Rawls of *PL* took up what he calls "the question of legitimacy" to show that a well-ordered society would be stable. Dreben thought that citizens' perception that laws are legitimate helps to stabilize the well-ordered society because their

---

<sup>25</sup> John Rawls, "Justice as Fairness: Political not Metaphysical", *Philosophy and Public Affairs* 14 (1985): 223-51, pp. 236-37, note 19.

<sup>26</sup> In my *Why Political Liberalism: On John Rawls's Political Turn* (Oxford University Press, 2010)

perception of legitimacy engages their sense of duty so that they see they are “politically-morally bound to obey the law”.<sup>27</sup>

I said earlier that the standard reading has much to be said for it. I shall not detail all the textual support that could be marshaled for the standard reading or that is marshaled for the interpretive comments about *PL* that I have said are underlain by that reading. I shall simply note some especially important passages in *PL* that seem to support the first three steps of the standard reading. If those steps are sound, then the conclusion expressed at the fourth step – about the subject matter of *PL* – is a natural conclusion to draw.

Rawls seems to say that he recast justice as fairness for just the reason alleged at the first step, for when he explains his transition to political liberalism, he says that in *TJ*, he had unrealistically assumed that all the members of the WOS would “endorse [justice as fairness] on the basis of what I now call a comprehensive philosophical doctrine”. (*PL*, p. xviii) Context makes clear that Rawls means he had assumed they would all endorse it on the basis of *the same* comprehensive doctrine. This text therefore lends some support to the standard reading’s contention that Rawls recast justice as fairness because he had assumed members of the WOS would all endorse the principles of justice on the basis of a shared comprehensive Kantianism. Rawls undoubtedly allows what is alleged at the second step: that members of the well-ordered society would disagree about justice as fairness even after it is recast as a political liberalism. In “Idea of an Overlapping Consensus”, he concedes that the focus of an

---

<sup>27</sup> Dreben, “Rawls and Political Liberalism”, p. 327.

overlapping consensus is likely to be, not a single conception of justice or a single set of principles, but “a class of liberal conceptions that vary within a certain more or less narrow range” (*PL*, p. 164). In the “Introduction to the Paperback Edition”, he grants that citizens of a well-ordered society might well endorse different liberal political conceptions of justice (*PL*, p. xlvi). If we accept Dreben’s explanation of the third step, then there is textual support for that step as well. For Rawls seems to say that his central concern in *PL* is just the concern which Dreben says led him to take the third step and introduce the principle of legitimacy – namely, the concern with stability. (*PL*, p. xvii)

The standard reading also seems to derive support from its ability to explain a change in Rawls’s view that many readers have thought accompanied the recasting of justice as fairness. To see this, recall Williams’s remark that in *PL*, Rawls “wants to make a bigger gap than *TJ* allowed between two different conceptions: that of a society in which power is rightfully exercised (a well-ordered society), and that of a society that meets liberals’ aspirations to social justice.” Rawls is said to have distinguished the two conceptions by shifting to a weaker standard of public justification and by characterizing a well-ordered society as a society in which power is legitimately exercised. According to the standard reading, Rawls distinguished those conceptions to solve the problem of stability. But the cost of distinguishing the conceptions and solving the problem was to open a gap between the two and to allow that the WOS of *PL* need not “meet[] liberals’ aspirations to social justice”. More specifically, it is said,

the WOS of *PL* need not satisfy the difference principle, which is the more controversial of Rawls's two principles.

The retreat from the difference principle that many readers think they have detected in *PL* is just what we would expect if we accept the standard reading's account of why Rawls recast justice as fairness and shifted his focus from justice to legitimacy. As if to confirm the naturalness of the expectation, some critics have alleged that Rawls retreated precisely because the difference principle would be controversial under conditions of pluralism and because controversy over fundamental principles threatens stability. Simone Chambers sums up an argument to this effect by saying that "Rawls's deep commitment to equality is in tension with his equally deep and democratic commitment consensus as the starting point for justice."<sup>28</sup> The tension in Rawls's view had to be resolved, she says, and "[t]he search for stability led Rawls to push controversial principles concerning social justice into the background and to place more widely accepted views concerning rights and freedoms into the foreground."<sup>29</sup>

Thus the standard reading seems to have, not only textual support, but enough explanatory power both to unify interpretive remarks by a number of careful readers of *PL* and to account for what some of them take to be a disturbing retrenchment in Rawls's later thought. In the next two sections, I shall argue that despite its power, the

---

<sup>28</sup> Simone Chambers, "The Politics of Equality: John Rawls on the Barricades", *Perspectives on Politics* 4,1 (2006): 81-89, p. 81; also her note 10 and the accompanying text. See also Eric Gregory, *Politics and the Order of Love* (University of Chicago Press, 2008), pp. 71-72.

<sup>29</sup> Chambers, "Politics of Equality," p. 81.

interpretation is mistaken. Another reading of *PL* is called for which assigns the liberal principle of legitimacy a different role.

*§2. Difficulties with the Standard Reading: The First Three Steps*

Let us start with the question of whether the Rawls of *PL* did indeed retreat from his earlier commitment to economic egalitarianism.<sup>30</sup> We can begin to consider that question by looking into the first step of the standard reading. As I indicated in the introduction, I believe that reading goes wrong at that step by misidentifying Rawls's reasons for recasting justice as fairness. I cannot definitively establish this claim here; I shall simply point to a serious difficulty with the standard reading of Rawls's political turn.

As I mentioned earlier, Rawls implies that in *TJ*, he had unrealistically assumed that all members of the well-ordered society would accept the same comprehensive view and would endorse justice as fairness on that basis (*PL*, pp. xvi-xvii). To see why he recast justice as fairness, we have to see where he relied on that assumption. According to the standard reading, the arguments for the principles of justice in the first part of *TJ* depend upon distinctively Kantian premises. Rawls relied on the unrealistic assumption when he assumed that everyone in the well-ordered society would endorse his principles on the basis of arguments that depend on those premises.

---

<sup>30</sup> Estlund argues convincingly that even if *PL* allows that a WOS need not satisfy the difference principle, this does not imply that Rawls came to think that the difference principle is not a demand of justice. According to Estlund, it merely implies that the Rawls of *PL* distinguished what he continued to think that justice demands and what he came to think that legitimacy allows.

The problem with this reading is that comprehensive views – whether full or partial -- are *views of the good*. And so what *TJ* must have unrealistically assumed is that all members of the well-ordered society would have the same – possibly partial -- conception of the good. Rawls does not introduce his account of goodness until the third part of *TJ* and, with the exception of claims about primary goods, premises about the good do not enter into the argument for the principles. It is therefore hard to see where in that argument Rawls could have relied on the unrealistic assumption or how the assumption would have helped establish the principles of justice.

Where Rawls seems much more likely to have relied on the assumption is in the third part of *TJ*. There he tries to show – crudely put – that one of the reasons justice as fairness would be stable is that its members would all judge that it is good to be just. Showing this would be much easier if Rawls could assume that the institutions of the well-ordered society would encourage some convergence in everyone’s conception of the good – that is, if they encouraged everyone to endorse the same partial conception of the good. As I shall indicate at the beginning of the next section, this is the assumption Rawls relied on. Coming to see how unrealistic the assumption was, Rawls recast justice as fairness to take fuller account of pluralism about the good than he had in *TJ* and to fix the problem he found in the third part of that book.

If my reading is correct, then Rawls’s reasons for presenting justice as fairness as a political liberalism is to show that justice as fairness – understood as including *both* of Rawls’s two principles – would be stable for the right reasons. Even if he later became interested in how stability bears on societies that do not meet *TJ*’s “aspirations

to social justice”, as the standard reading says, showing the stability of justice as fairness *as formulated in TJ* was Rawls’s original reason for recasting it as a political liberalism. In that case, then not only is the standard reading wrong at the first step about what motivated Rawls’s political turn, but that turn cannot constitute a retreat from the difference principle.

Even if we grant proponents of the standard reading their first step, the reading may seem to face a serious textual difficulty. As I mentioned in the introduction, Rawls came to the realization referred to the second step – the realization that justice as fairness would be the object some disagreement even after it was recast as a political liberalism – only after the project of recasting it was well under way. Thus Rawls does not say anything about such disagreements in “Justice as Fairness: Political not Metaphysical”, the essay in which he began to present justice as fairness as a political conception. His remark about the likely focus of an overlapping consensus, which was one of two passages that I said supports the second step, occurs in the version of “Overlapping Consensus” that is included in *PL*. There is no comparable passage in the original version of that essay, which appeared some years before.<sup>31</sup>

The other passage that I said supports the second step of the standard reading is a similar concession from the “Introduction to the Paperback Edition”. But in the revised Dewey Lectures – which were included in *PL* but which antedate the “Introduction to the Paperback Edition” and which may antedate the revision of “Overlapping Consensus” – Rawls still repeats *TJ*’s description of the WOS virtually

---

<sup>31</sup> John Rawls, “Idea of an Overlapping Consensus”, *Oxford Journal of Legal Studies* 7 (1987): 1-25.

word-for-word, saying that a well-ordered society is one in which “everyone accepts, and knows that everyone else accepts, the *very same* principles of justice”. (*PL*, p. 35, emphasis added) Since the revised Deweys made their first appearance in *PL*, Rawls seems to have been committed to unanimity about justice well into the drafting of *PL*. The standard reading is therefore committed to the claim that *PL* is really about a concept – namely, legitimacy – that Rawls introduced to solve a problem he only became aware of well after he began to recast justice as fairness.

Defenders of the reading may reply that what really concerned Rawls in his later work was the possibility of showing how a WOS could be stable under conditions of pluralism. The story of his work after *TJ* is one of a deeply reflective thinker who came to an ever more profound appreciation for pluralism’s reach and power. Rawls first became aware of the problem pluralism posed for justice as fairness as he had presented it in *TJ*, as the first step of the standard reading says. He began to recast justice as fairness to address that problem. But as he thought about pluralism more deeply, he came to see that the same factors which threatened consensus on justice as fairness as originally presented also threatened consensus on it even when it was recast as a political conception, as the second step says. The third step can then be taken to say that Rawls introduced legitimacy to address the problem he finally came to appreciate. The claim at the fourth step -- that *PL* is really about legitimacy – is to be taken, not as a claim about what concept was central to all of the work Rawls collected in that book, but as a claim about what was central to the most mature presentation of his view found in the latest essays.

One problem with this reply is that if it is correct, then *PL* is not about what Rawls explicitly said it is about. For Rawls says quite clearly that the purpose of *PL* is to fix a problem with *TJ*. (See *PL*, pp. xvff. and xliff.) According to the standard reading, the problem with *TJ* is posed by the possibility of disagreement asserted at the first step, the possibility of disagreement about justice as fairness as originally presented. The standard reading interprets *PL* as addressing the problem posed by the possibility of disagreement asserted at the second step rather than the first. The two problems are different; what unites them, according to the defense of the standard view that I have just imagined, is that both grow out of the fact of pluralism.

Far from salvaging the situation, this last claim -- that pluralism of comprehensive doctrines and disagreement about political conceptions of justice can be traced to a common source -- raises further difficulties for the standard reading. Rawls is sometimes read as saying that human beings naturally differ about deep questions of all kinds. Those differences are manifest in free societies simply because, instead of repressing them, free societies tolerate them by honoring various liberties, such as the liberty of conscience, and freedom of speech, press and association. In the well-ordered society, it might be thought, human divergence plus equal liberty yield both pluralism about comprehensive views of the good (alleged at the first step of the standard reading) and pluralism about political conceptions of justice (alleged at the second).

But this genealogy of pluralism is too superficial. I cannot go deeply into Rawls's analysis of pluralism here, but very briefly: pluralism arises under free institutions because under those institutions we acquire, and we act under, an idea of

ourselves as free. The acquisition of this view of ourselves is part of the educative effect of full publicity. (*PL*, p. 71) The crucial question is what members of a just society would think of themselves as free to do. Rawls offers a clear answer: they would think of themselves as free to form, pursue and revise their conceptions of the good. (*PL*, pp. 30ff.) It is in part by encouraging this conception that free institutions encourage pluralism about comprehensive doctrines.

By contrast, institutions of a liberal society do not encourage their members to think of themselves as similarly free to form and revise their own conceptions of the right.<sup>32</sup> Indeed, they cannot. For if they did, they would -- in effect -- encourage their members to rethink the liberties that allow them the freedom to pursue conceptions of the good and to rethink the conception of themselves and their citizenship that justifies those liberties.<sup>33</sup> Thus it is only by encouraging convergence, or perhaps *bounded* disagreement, about the right that just societies can encourage pluralism about the good. It is therefore far from clear that disagreements about justice arise for the same reasons as pluralism about the good. Perhaps they do. Or perhaps different views of the good have different conceptions of justice associated with them, so that differences about the good give rise to disagreements about justice.<sup>34</sup> Or perhaps reasonable pluralism about the good is compatible with unanimous agreement on a conception of justice.

---

<sup>32</sup> As Rawls observes in an important argument in *TJ*, “the good of individuals” may be “up to them to decide”; “what is right”, by contrast, “is not a matter of mere preference”. (*TJ*, p. 490)

<sup>33</sup> I am grateful to Ben Laurence for helpful conversation about this point.

<sup>34</sup> Jeremy Waldon distinguishes these cases in *Law and Disagreement* (Oxford University Press, 1999), pp. 149ff.

One indication that Rawls thinks the two kinds of disagreement are fundamentally different is that in a WOS, pluralism about the good would range much more widely than pluralism about justice. For though the comprehensive doctrines adhered to in a well-ordered society would largely be reasonable, their being so is compatible with wide variation among doctrines all of which are on a par from a political point view. By contrast, the conceptions of justice that would be endorsed in a WOS would “vary within a certain more or less narrow range” (*PL*, p. 164) and would have a “focal class” (*PL*, p. 174). But if the pluralism alleged at the first step of the standard reading and that alleged at the second are significantly different – and, in particular, if they arise in different ways -- then it is at best misleading to say that the standard reading captures the result of Rawls’s deepening insight into a single phenomenon called “pluralism”.<sup>35</sup>

What of the third step in the standard reading, the claim that the possibility of disagreement about the principles as presented in *PL* led Rawls to shift the focus of his attention from justice to legitimacy? Doubts about this claim are raised by what Dreben says legitimacy is for. He writes:

To say that a law is a legitimate law is not to say that all reasonable citizens agree with it. ... What you really have to worry about in a liberal constitutional democracy is how, when a law is appropriately passed, it is binding on all citizens, even on those citizens who reasonably can differ with it.

Dreben correctly identifies “what you really have to worry about”, or one of the things “you really have to worry about” in a liberal democracy. And he is right that Rawls’s

---

<sup>35</sup> Contrast Reidy, “Reciprocity and Reasonable Disagreement”, p. 247.

treatment of legitimacy shows how that worry is to be addressed. The problem for the third step is that the worry Dreben identifies does not depend upon disagreements of the sort alleged at the second step – disagreements at the level of Rawls’s two principles. For even if citizens of a WOS were unanimous in their agreement on the principles, they could still disagree about the justice of laws their society enacts.

Rawls considers this possibility in *TJ*, where he notes that citizens who agree to the two principles could still disagree about economic policies enacted at the legislative stage. (*TJ*, pp. 174ff.) In *PL* and later writings, he considers Catholics who regard a liberal abortion regime as unjust and Quakers who find their country’s military policies unjust. These two cases are relevantly like those Rawls considers in *TJ*: they are disagreements about the justice of law and policy that could arise among citizens who all agree with the principles of justice.<sup>36</sup> As I said at the outset, and as I shall try to document below, *PL*’s treatment of legitimacy does not add anything new to what *TJ* said about why the law is binding in such cases. And so, contrary to the third step of the standard reading, I do not think Rawls introduced liberal principle of legitimacy to make a new point about the duties of citizens in light of disagreements of the kind alleged at second step. Rather, as well shall see, he introduced it to stress and apply a point he had already made, and to make a new point about citizens’ good.

---

<sup>36</sup> It is hard to see how the two principles bear at all on the policies to which the Quakers are said to object, so their objections must be independent of their view about the principles and therefore compatible with their acceptance of them. As for abortion: Rawls quite clearly implies that disagreements about abortion policy can arise from disagreement, not about what fundamental political values are, but about how those values are to be balanced (*PL*, p.243 note 32).

§3. *Difficulties with the Standard Reading: The Project of Political Liberalism*

The conclusions I have reached so far about the first three steps of standard reading raise serious doubts about the fourth step, its contention that legitimacy is what *PL* is about. But there are also ample independent grounds for doubting that contention.

If *PL* were about legitimacy in the same sense of ‘about’ in which *TJ* is about justice, we would expect Rawls to give a clear statement of the concept or meaning of the term and to give some systematic treatment of its various conceptions, ideally indicating why his favored conception would be adopted in the original position. Instead, what Rawls has to say about legitimacy is maddeningly brief and vague. As John Simmons notes, he never says what ‘legitimacy’ actually means.<sup>37</sup> Rawls does say that the principle of legitimacy he favors would be adopted in the original position, but he says nothing about how the parties in the original position make their decision.

Even more problematic for the standard reading is the fact that Rawls fails to engage central questions about legitimacy, including questions whose centrality he quite clearly suggests. For example, late in *PL* Rawls says of legitimacy that “reasonable citizens understand this idea to apply to the general structure of authority” (*PL*, p. 393). On one way of taking this remark, legitimacy attaches to the political structure – i.e. to that part of society’s basic structure -- which exercises authority. On this way of taking it, Rawls thinks that legitimacy applies, at least in the first instance, to a society’s governing apparatus, the state.

---

<sup>37</sup> A. John Simmons, “Justification and Legitimacy”, in *Justification and Legitimacy: Essays on Rights and Obligations* (Cambridge University Press, 2001), pp. 122-57, p. 145, note 48.

There are many understandings of state legitimacy on offer in contemporary political philosophy. State legitimacy is sometimes thought of as the state's possession of a permission right or a justification right. On this understanding, if a state or a government is legitimate, then it is permitted to or is within its rights in doing what states characteristically do – issue directives, employ force and enforce a monopoly on ultimate coercive power. Sometimes, the legitimacy of a state or government is thought of as being or entailing the power to change the moral status of subjects by imposing obligations which it can expect them to obey. On this understanding, legitimacy confers a claim-right, since legitimate states and governments can claim obedience. Sometimes legitimacy is said to be, as Simmons says, a “complex right” to impose obligations, claim obedience and coerce the non-compliant.<sup>38</sup>

Each of these ways of understanding legitimacy answers to some of our intuitions. One central question about legitimacy is which if any of these understandings is best. Unfortunately Rawls's liberal principle of legitimacy, and associated remarks, shed surprisingly little light on that question. The principle says:

Our exercise of political power is proper and hence justifiable only when it is exercised in accordance with a constitution the essentials of which all citizens may reasonably be expected to endorse in the light of principles and ideals acceptable to them as reasonable and rational. (*PL*, p. 217)

This principle does not seem to supply a standard of *state* legitimacy at all, since it seems to apply to exercises of political power rather than to the “structure” or state which exercises that power. Moreover, what it suggests about the various

---

<sup>38</sup>See Simmons, “Justification and Legitimacy”, p. 130.

understandings of state legitimacy I canvassed just above is somewhat confusing, at least if taken in conjunction with some of the other things Rawls says. The statement of the principle – and in particular, the phrase “proper and hence justifiable” -- suggests that Rawls thinks a state which exercises power legitimately thereby exercises a justification right.<sup>39</sup> Elsewhere, however, Rawls says that a legitimate political authority can issue laws that are “politically (morally) binding”<sup>40</sup>; he also implies that citizens have a duty “not to violate” legitimate law and “not to resist [it] with force”.<sup>41</sup>

These last passages suggest that the legitimate exercise of power is the exercise, not just of a justification right, but of a claim right. The standard reading must place a great deal of weight on these passages, since that step says that citizens’ perception of legitimacy stabilizes by engaging their sense of duty. Rawls’s remark that legitimate laws are “politically (morally) binding” may seem to justify the third step, for Dreben seems merely to repeat the remark when he says legitimacy stabilizes because citizens who see that a law is legitimate thereby see that they are “politically-morally bound to obey” it.<sup>42</sup> But if legitimacy were as central to *PL* as the standard reading contends at its fourth step, and central for the reason asserted at the third step, we would expect Rawls to argue that this last understanding of legitimacy is the right one and to develop

---

<sup>39</sup> See Simmons, “Justification and Legitimacy”, p. 145, note 48; also Allen Buchanan, “Recognitional Legitimacy and the State System”, *Philosophy and Public Affairs* 28 (1999): 46-78, p. 57.

<sup>40</sup> John Rawls, *Collected Papers* (Harvard University Press, 1999) ed. Freeman, p. 578. This work will hereafter be cited as ‘*CP*’.

<sup>41</sup> *CP*, pp. 594-95 note 57, and 606.

<sup>42</sup> Dreben, “Rawls and Political Liberalism”, p. 327.

his remarks about the duties of compliance and non-violent resistance. Instead these remarks are left as suggestive asides.

Thus Rawls's treatment of legitimacy leaves enough questions unanswered, and suggests enough different views about legitimacy, that we should be suspicious of the standard reading's claim that legitimacy forms the subject matter of *PL*. In the next section, I shall sketch a reading of *PL* that implies a very different role for the liberal principle of legitimacy. According to the reading I shall suggest, *PL* is a sustained attempt to fix justice as fairness so that it avoids problems Rawls found in a clearly identifiable set of arguments in *TJ*. Rawls developed the defining concepts of political liberalism -- such as liberal legitimacy, public reason and overlapping consensus -- in order to effect the repairs. But *PL* is no more about one of those concepts than it is about the others. To conclude that it is about any of them, and to reach that conclusion by taking the first three steps in the standard reading, obscures the repair effort which is the central project of that book. To see that, we have to see why Rawls came to think repairs were necessary.

#### *§4. Stability in Political Liberalism*

As I noted earlier, Rawls says that he recast justice as fairness because of an inconsistency in *TJ*'s account of stability. (*PL*, p. xviii) I have argued that the standard reading mistakes the source of the inconsistency and misidentifies the reasons for Rawls's political turn. To see where the inconsistency lies, it is useful to recall how the argument for stability goes.

*TJ*'s argument for stability depended upon two conclusions. Rawls had argued that justice as fairness would be stable first, because members of a WOS would all acquire a sense of justice informed by the principles of justice and second, they would all affirm that maintaining their sense of justice belongs to their good. (Cf. *TJ*, p. 450) In *TJ*, Rawls refers to the second conclusion as the “congruence” of the right and the good. Rawls remained satisfied with the argument he had offered for the first stability conclusion in *TJ*, chapter 8<sup>43</sup> and he continued to rely on a similar argument in *PL*. (*PL*, p. 141) What he came to see was that *TJ* had relied on unrealistic assumptions about the good in order to reach the second stability conclusion in chapter 9.

More specifically: to show congruence, Rawls assumed that a WOS would encourage members' views of the good to converge, so that – wherever they differed – they all valued their sense of justice for the same reasons, based on the same ethical values and ideals, and on the same desires for certain goods whose value was accounted for by the thin theory of goodness. (cf. *TJ*, pp. 498ff.) It was by assuming this partial convergence that Rawls supposed members of well-ordered society would share a “comprehensive, or partially comprehensive, doctrine[.]” (*PL*, p. xviii) But, Rawls came to think, the assumption that just institutions would encourage such convergence is unrealistic and is inconsistent with the fact that those institutions encourage pluralism about the good. Rawls made the transition to *Political Liberalism* – and introduced a set of distinctively political values and ideals – so that he could argue for the second

---

<sup>43</sup> Rawls, *Justice as Fairness: A Restatement* (Harvard University Press, 2001) ed. Kelly, p. 196, note 17. I do not think Rawls ever reconciled his expressed satisfaction with his late recognition that members of the well-ordered society would differ fundamentally about justice. While I think materials for a reconciliation can be found in Rawls's writings, I cannot pursue the matter here.

stability conclusion without relying on an unrealistic convergence, and to answer questions raised by the new argument.<sup>44</sup>

Here is a very rough approximation of the new argument for the second stability conclusion:

- (1) In a WOS, justice as fairness would be supported by an overlapping consensus of reasonable comprehensive doctrines.

When an overlapping consensus obtains, “reasonable doctrines endorse the political conception, each from its own point of view.” (*PL*, p. 134) So:

- (2) According to each reasonable comprehensive doctrine, the good of realizing the political ideals, values and principles of justice as fairness “normally outweigh whatever values are likely to conflict with them”. (*PL*, p. 156)

What happens, then, when reasonable members of the well-ordered society take up the viewpoint of their comprehensive doctrines and assess justice as fairness? If an overlapping consensus obtains, as (1) says, and if (2) follows from (1), then:

- (3) Reasonable members of the WOS all affirm that the good of acting from the values and principles of justice as fairness, and of living up to its ideals, “normally outweigh whatever values are likely to conflict with them.”<sup>45</sup>

If that’s right, and if members of the WOS anticipate confronting only normal political circumstances, then:

---

<sup>44</sup> I argue for this in my *Why Political Liberalism*.

<sup>45</sup> Thus in “Reply to Habermas”, Rawls says that when an overlapping consensus obtains and is known to obtain: “citizens will judge (by their comprehensive view) that political values either outweigh or are normally (though not always) ordered prior to whatever nonpolitical values may conflict with them.” (*PL*, p. 392)

(4) Reasonable members of the WOS all affirm that maintaining an effective desire to act from the values and principles of justice as fairness, and to live up to its ideals, belongs to their good.

Since the desire referred to in (4) just is a sense of justice, Rawls can get to the conclusion he wants:

C: Reasonable members of the WOS all affirm that maintaining an effective sense of justice, informed by justice as fairness, belongs to their good.

If members of the WOS would all acquire a sense of justice, and if the argument for C is sound, then members of the WOS will develop and affirm allegiance to a just constitution and will support the institutions that implement it. Then justice as fairness will be stable for the right reasons.

My statement of the argument for C is regrettably rough. Many details need to be filled in, and game-theoretic problems about assurance have to be overcome.<sup>46</sup> A full defense of the argument would show that it is more realistic to suppose that just institutions would encourage an overlapping consensus on political values, as Rawls does in the first premise, than it is to suppose that they would encourage the convergence on ethical values and rational desires that Rawls had assumed in *TJ*. I pass over these questions because my primary concern here is to recover an account of legitimacy that fits into this reading of Rawls's project in *PL*. I provide just enough details about that central argument to motivate the account of legitimacy.

---

<sup>46</sup> When the assurance problems are overcome, "[t]he plan of life which [affirms one's sense of justice] is his best reply to the similar plans of his associates." (*TJ*, p. 497) In that case, the well-ordered society satisfies the condition of a Nash equilibrium.

Note that at crucial steps – namely, (2) and (3) – the new stability argument depends upon political values and ideals of justice as fairness being seen to outweigh values and ideals that compete with them. I believe that among the political ideals of justice as fairness as laid out in *PL* is one that we might call the *Ideal of Democratic Governance*. This is “the ideal of citizens governing themselves in ways that each thinks the others might reasonably be expected to accept”. (*PL*, p. 218; cf. pp. 139-40) This ideal is realized in the on-going political life of a well-ordered society when citizens follow the guidelines of public reason and govern themselves in accordance with a just constitution.

Though I cannot go into detail here, I believe that the political *Ideal of Democratic Governance* is introduced to replace an ethical ideal that did similar work in *TJ*, but that depended upon more convergence on the good than it is reasonable to expect in a pluralistic society like the well-ordered society. Very briefly: the second stability conclusion – that members of a WOS would affirm that maintaining their sense of justice belongs to their good -- depends upon the claim that they would experience political society itself as a good. In *TJ*, Rawls had argued that it would be experienced as a good because it would realize the ideal of a social union of social unions. (*TJ*, §79) He came to realize that the social union of social unions was “no longer viable as a political ideal once we recognize the fact of reasonable pluralism.” (*PL*, p. 388 note 21) The stability argument of *PL* therefore required a different argument that political society would be experienced as a good, an argument that appealed to a different and less demanding political ideal. That ideal is the *Ideal of Democratic Governance*.

To see how the *Ideal of Democratic Governance* contributes to stability, consider:

A constitution specifies the structure of government. It says of what organs government consists and how power is apportioned among them. A constitution is also, as Rawls in *TJ*, “the highest-order system of social rules for making rules”. (*TJ*, p. 195) As such, it specifies the procedures by which policies are made, laws are enacted and the constitution itself is amended. It also specifies the rights and liberties which must be respected when government power is exercised. These procedures, rights and liberties are the “essentials” of the constitution. (*PL*, pp. 227ff.) If the procedures are familiar democratic ones for arriving at decisions when unanimity cannot be expected, and if the rights and liberties are given by a principle chosen in the original position, then citizens of the WOS may reasonably be expected to endorse the essentials of their constitution.

Rawls says at one point that “[t]he aim of public reason is to articulate” an ideal expressed by a democratic constitution: “the ideal of a people to govern itself in a [democratic] way”. (*PL*, p. 232) Whatever else he means by this, Rawls clearly thinks that those exercising political power under a democratic constitution – including voters (*PL*, p. 217) -- should adhere to the guidelines of public reason. Thus on-going governance in accord with a democratic constitution realizes many very great political values, including the protection of individual liberty, the provision of a social minimum, the rule of law, transparency, mutual respect, political autonomy and adjudication of central questions by clearly known public procedures on the basis of values that all can

accept. Because justice as fairness shows how these values fit together, it is accurate to say that the well-ordered society of justice as fairness realizes a unified political ideal of self-government.

Seeing that this ideal – and its very great constituent goods -- is available when citizens govern themselves in accordance with a just democratic constitution, we can begin to see why Rawls thinks citizens who adhere to a variety of comprehensive doctrines would find some intrinsic value in a just constitutional democracy. And so we can begin to see why Rawls thinks that C and its consequences are true. That is, we can begin to see why the members of a WOS would affirm that their own sense of justice and their own support for the constitution are traits of character that belong to their good. We could begin to see why they would affirm that those are traits they will try to reinforce rather than to uproot, so that they can participate in the political ideals that a WOS makes available.

But if the argument for C is right, then they must reach these judgments because – as (3) says -- they judge that realizing the *Ideal of Democratic Governance* normally outweighs, or normally helps to outweigh, competing values. And – as the transition from (3) to (4) requires -- they must anticipate confronting only normal political circumstances. On my reading, Rawls appeals to the notion of legitimacy to *how* that value can be seen to outweigh competing ones in the face of difficult cases, cases which might seem abnormal enough that some citizens are tempted to rethink their commitment to justice. Showing this will show how (3) can be true and how the new stability argument can succeed and the inconsistency Rawls found in *TJ* can be avoided.

*§5. The Difficult Cases*

What are the difficult cases? In *TJ*, Rawls imagines a multi-stage sequence for the choice and implementation of principles of justice. (*TJ*, §31) The kind of cases I have in mind arise at the stages at which the principles are implemented. Though I believe difficult cases can arise at the constitutional stage, the stage at which the constitution is written, I shall ignore that complication. Instead, I shall suppose that members of the WOS value the ideals realized under a just constitution, but that their attachment to the constitution can be threatened at later stages of implementation.

I have suggested that Rawls made the transition to *PL* because he came to see that the argument for C could fail for some members of the WOS. It is sometimes alleged – as by Stephen Holmes<sup>47</sup> -- that Rawls made the transition from *TJ* to *PL* because he came to think that religious believers might reject justice as fairness as presented in *TJ*. And so it would be natural to elaborate my suggestion by saying that those of whom the argument for C fails are those Rawls calls “citizens of faith”.

While I believe that *TJ*'s arguments for C can be shown to fail for some reasonable religious believers in the well-ordered society, I do not think that that fact is what actually led Rawls to recast justice as fairness as a political liberalism. I do think, though, that as Rawls reformulated his view, he took up some of the problems he did because of the likelihood that the WOS would include reasonable citizens of faith whose religious views shape their views about the justice of legislation. This includes

---

<sup>47</sup> Stephen Holmes, “The Gatekeeper: John Rawls and the Limits of Toleration”, *The New Republic*, October 11, 1993, pp. 39-47. See also Dreben, “Rawls and Political Liberalism”, p. 345.

the problem posed by the hard cases legitimacy is supposed to address. So consider some member of the WOS Jan, whose religious views ground objections to an exercise of legislative power.

Jan may, like the Catholics to whom Rawls refers in “Public Reason Revisited”, have religious objections to the legislative enactment of the abortion regime Rawls suggests that justice allows. Or she may be a religious pacifist who opposes the levying of taxes to support the use of military power sanctioned by *Law of Peoples*, like the Quakers Rawls referred to “Reply to Habermas”.<sup>48</sup> (*PL*, pp. 393-94) In these cases, I am supposing Jan judges that the use of political power to implement the principles is unjust according to her religion’s standards of justice. By her lights, there are important goods to be realized in political life – such as such as those she thinks would be realized by pacifism or by the legal protection of fetal life – that are not realized in the well-ordered society, but that she thinks would be realized with different legislation. Moreover, by exercising political power as it does, she thinks, the WOS allows serious injustices. The combination of opportunity cost and injustice may lead Jan to have serious reservations about the constitution that implements justice as fairness.

That Jan has these reservations does not in itself tell against step (3) of the stability argument. In fact, the wording of step (3) suggests that Rawls anticipated that some members of the WOS would have such reservations. (3) says that reasonable

---

<sup>48</sup> I believe Rawls thinks difficult cases can arise at the constitutional stage as well as at later stages because of what he says raise the question of legitimacy. He says that question is raised by the imposition of power on “citizens some of whom may not accept the reasons widely said to justify the structure of political authority – the constitution – or when they do accept that structure, they may not regard as justified many of the statutes enacted by the legislature to which they are subject.” (*PL*, p. 136)

members of the WOS all affirm that acting from the values and principles of justice as fairness, and living up to its ideals, “normally outweigh whatever values are likely to conflict with them”. If we include the sum of injustice and opportunity cost in what has to be outweighed, then (3) seems to presuppose conflicts of value of just the sort that Jan sees. So long as there are political values which Jan thinks normally tip the balance in favor of justice as fairness, (3) will be true of her. The question is what those values are and how much weight Jan is likely to attach to them. I have said that one of the values to which (3) refers is the value of realizing the *Ideal of Democratic Governance*. Can that value tip Jan’s balance of reasons in favor of a constitution which implements justice as fairness but is thought to yield an unjust law?

It’s not at all clear that it can. For Jan may deny that the *Ideal of Democratic Governance* is realized in the problem case. Suppose she knows enough democratic theory to think of political power as Rawls does -- as “the power of the public, that is, the power of free and equal citizens as a collective body” (*PL*, p. 136) of which each citizen has “an equal share”. (*PL*, p. 217) So she thinks that political power is in part hers and that its exercise must be justifiable to her, particularly if that power – her power -- is said to be exercised on her behalf. Suppose further that Jan thinks exercises of power are justifiable only if they satisfy the justificatory standard which is spelled out at length in *TJ*: she thinks they are justifiable only if they are just. Then, because she thinks political outcomes in the cases I have imagined are unjust, she will think power has been exercised in a way that cannot be justified to her and that an important requirement of political morality has not been met. Moreover, Jan will think, others

should know that people with her beliefs will find such exercises of power unjust and hence unjustifiable, and so they could not have expected her to endorse them. And so she will think that the good of *Democratic Governance* has not been realized in the difficult cases, and not realized precisely because of what makes those cases difficult in the first place: they are cases in which, from Jan's point of view, an unjust law has been enacted.<sup>49</sup>

Of course, even this conclusion does not show that (3) is false of Jan or that Jan thinks the legislation to which she objects is an abnormal case that leads her to rethink her commitment to a just constitution. There may be other political goods that do not depend upon the justice of political outcomes and that Jan thinks *do* outweigh the conflicting values in the cases I have imagined. But since I want to make a point about the connection between legitimacy and the *Ideal of Democratic Governance*, I shall – somewhat artificially -- ignore this possibility. Or Jan may think that since there are so many other cases in which the good of *Democratic Governance* is available, her balance finally tips in favor of justice as fairness and the constitution, despite the reservations engendered by the difficult cases.

But now suppose that Jan judges one or another of the difficult cases to be extremely important, and the combination of opportunity cost and injustice in those cases to be very weighty.<sup>50</sup> The fact that the good of *Democratic Governance* is realized

---

<sup>49</sup> An interesting example is the organization “Not in Our Name” created by opponents of the Iraq war; see <http://www.notinourname.net/>.

<sup>50</sup> In fact, the cases Rawls picks – of Catholics who oppose abortion and Quakers who oppose war (*PL*, pp. 393-94) – are presumably picked precisely because they illustrate this possibility.

in other cases may not seem to her to outweigh the injustice and opportunity cost in these cases, especially since she thinks that good is not realized in *them*. Then the case or cases *will* pose difficulties for (3) or for the move from (3) to (4). Because of the forces of social learning at work in a well-ordered society, Jan may have some tendency to judge and act according to justice as fairness. But if she does, she will be of two minds about justice and may think, on reflection, that that tendency misleads her in important cases. (4) would then be false of her, and Rawls could not infer what I have said is his desired conclusion C.

I am not supposing that the failure of the argument for C implies that Jan and those like her will rebel against the WOS or will otherwise attempt to destabilize it. But I am entertaining the possibility that they will live in the WOS in much the same way that religious ethicist Stanley Hauerwas says Christians like himself live in contemporary America: as “resident aliens” whose status is reflected, not in their reduced legal privileges, but in their alienation from or their ambivalence about liberal values.<sup>51</sup> And so while the presence of citizens like Jan would not show that the WOS would be unstable or unjust, it would force us to ask whether justice as fairness, and the constitution which implements it, would be stable for the right reasons. It would therefore raise troubling questions about *PL*’s treatment of stability.

Lest we doubt that these questions troubled Rawls, note that he says one of the most important questions confronting political liberalism is:

---

<sup>51</sup> See Stanley Hauerwas and William H. Willimon, *Resident Aliens: Life in the Christian Colony* (Abingdon Press, 1989).

How is it possible for citizens of faith to be wholehearted members of a democratic society when they endorse an institutional structure satisfying a liberal political conception of justice with its own intrinsic political ideals and values, and when they are not simply going along with it in view of the balance of political and social forces? (*PL*, p. xxxviii)

If we attach some emphasis to ‘wholehearted’, as I am inclined to do, then we can see that the objection I have used Jan to develop is one Rawls took quite seriously. We gain considerable insight into Rawls’s reasons for introducing legitimacy by reflecting on how he would reply to it. How would Rawls respond, so that *PL*’s stability argument succeeds?

*§6. Stability and Legitimacy in Political Liberalism*

I have supposed that Jan judges certain exercises of power to be unjust on the basis of her religious comprehensive doctrine. Rawls could argue that comprehensive doctrines like Jan’s will not be found in a WOS because the social forces at work in that society will liberalize comprehensive doctrines, at least in the longer run – where ‘liberalize’ means

bring it about that either comprehensive views do not imply the injustice of political outcomes permitted by justice as fairness or, if they do, they do not attach sufficient weight to cases of injustice to falsify (1), (2), (3) and (4).

Thus, he might say, the sociology of the WOS insures the success of the stability argument.

There are scattered remarks which suggest that Rawls thinks such liberalization will occur in the WOS. In “Idea of Public Reason”, for example, he says that “[a] reasonable and effective political conception may bend comprehensive doctrines toward

itself[.]’ (*PL*, p. 246) Part of what Rawls has in mind, I think, is that when citizens of diverse comprehensive views live under a constitution which implements justice as fairness, they may come to appreciate the “intrinsic political ideals and values” of that conception of justice, come to value self-government under the constitution for its own sake and come to do so from within their comprehensive doctrines. But the ‘may’ in Rawls’s remark also betrays a tentativeness that is appropriate in the present connection, since the answer I am now considering to the difficult cases would require a lot of bending. Liberal democratic institutions are of relatively recent provenance, at least compared to the major world religions. At this point, we do not know enough about the long-term effects of liberal democratic institutions on religion to know whether any more definite an assertion is warranted. Moreover, such historical evidence as we have bears on the effects of institutions in societies that are far from well-ordered. So it would be a mistake for Rawls to rely exclusively on this strategy for dealing with Jan’s case and salvaging *PL*’s stability argument.

Rawls could hold that Jan’s religious view is false, at least as regards the injustice of the constitution or the legislation in question. He could also hold that the state may proceed on the assumption that her view is false, and may try to bring it about that Jan and her co-religionists recognize its falsity. Jan’s judgment that the constitution or legislation is unjustified rests on the prior judgment that it is unjust. So the state’s attempt to persuade her of the falsity of her religious views would, if successful, have the effect of disabusing her of the judgment that threatens the truth of (3), the move to (4) and the soundness of the stability argument. For good and familiar reasons,

however, Rawls is reluctant to pronounce on the truth of comprehensive doctrines. (*PL*, p. 127) I shall not review his reasons here. For now, suffice it to say that Rawls does not appeal to the falsity of Jan's religious views to answer the challenge she poses.

Rawls could maintain that Jan is being unreasonable and that she is therefore not a counterexample to propositions (3) and (4), which refer to reasonable citizens. This is a strategy Rawls is often accused of relying on to dismiss difficult cases, especially cases involving those who oppose abortion.<sup>52</sup> But, as I believe Rawls himself actually recognized, it will not do. We have already seen that Rawls cannot count on the social forces of a WOS liberalizing comprehensive doctrines enough to bring it about that there are no citizens like Jan. Indeed, for all Rawls has shown, there may be a large number like her in the WOS. If there are and if they are deemed unreasonable, then the stability argument laid out above would not show what it is supposed to show. For the argument is not intended to show how some subset of citizens in a well-ordered society can support justice as fairness for the right reasons, while the rest regard it as a *modus vivendi* or "go[] along with it in view of the balance of political and social forces". (*PL*, p. xl) The argument is supposed to show how justice as fairness can enjoy support for the right reasons *society-wide*. The argument will show *that* only if (3), (4) and C all refer to the overwhelming majority of citizens – hence only if an overwhelming majority of the citizens are taken to be reasonable. So, while dismissing Jan and citizens like her as unreasonable may salvage the stability argument in name, it would do so in name *only*, by sacrificing the larger point that Rawls wants to make.

---

<sup>52</sup> Michael W. McConnell, "Review: Religion and the Search for a Principled Middle Ground on Abortion", *Michigan Law Review* 92, 6 (1994): 1893-1909; see, in particular, p. 1893 note 3.

If *PL*'s treatment of stability is to succeed, then, Rawls must take Jan to be reasonable and he must show how (3), (4) and C can be true of her. That means he must show how Jan can take the values, principles and ideals of justice as fairness to outweigh competing values even in difficult cases. As we have seen, he must show how she can do so without denying the truth of her religious views.

At this point, it will help to recall why Jan posed a difficulty for the stability argument of *PL*: because of her religious views, Jan thinks that in an important case or cases, the WOS fails to realize certain great goods in political life and it enacts serious injustices. Step (3) suggests that Jan will take the great political goods that are realized in a WOS normally to outweigh this combination of injustice and opportunity cost. The problem with this suggestion is that Jan does not have the good of *Democratic Governance* available in the difficult or abnormal cases to do the outweighing. For that good is available to Jan only if she see that the relevant exercises of power are – and are believed by others to be -- justifiable to her. She thinks they are not justifiable to her precisely because she thinks those exercises of power issue in unjust outcomes and because she thinks the justice of outcomes is necessary to justify the exercises of power that produced them<sup>53</sup>.

Thus the problem Jan poses for the stability argument of *PL* is a compatibility problem: Jan's religiously-informed beliefs about justice are incompatible with the belief that the exercises of power in the difficult cases are justifiable. To avoid the problem, Rawls needs to show how these beliefs can be compatible after all. The Rawls

---

<sup>53</sup> More plausibly: she takes the justice of intended outcomes to be necessary to justify the exercises of power that produced them.

of *TJ* provided one salient standard of political justifiability: justice. The Rawls of *PL* could show how the two beliefs can be compatible if he could identify and render salient some other standard of justifiability that applies to exercises of political power, one that is less demanding than justice and that it is satisfied in the difficult cases. The liberal principle of legitimacy supplies that standard.

According to the standard reading, the difficult cases arise because of disagreements about fundamental principles of justice. The question they raise is that of how citizens like Jan can come to recognize their duty not to use force in resisting unjust laws.<sup>54</sup> Rawls is said to have introduced the liberal principle of legitimacy to answer that question and show stability. The problem with this reading is that the difficult cases concern disagreement about the justice of laws and policies. As I argued earlier, such cases can arise even without disagreement about principles of justice. The standard reading is correct to claim that Rawls thinks Jan is duty-bound not to resist laws with force even if she thinks they are unjust, and that citizens' recognition of this duty contributes to the stability of justice as fairness and the constitution that implements it. But we shall see that *PL* does not add to *TJ*'s treatment of the duty to obey the law. By supposing that the only question raised by the difficult cases concerns citizens' sense of right, and by exaggerating the novelty of *PL*'s answer to that question, the standard reading obscures what really is new and central to *PL*: the new stability argument sketched at the beginning of §4. The liberal principle of legitimacy helps to address questions that the difficult cases raise about that argument.

---

<sup>54</sup> See Dreben, "Rawls and Political Liberalism", p. 327.

As we saw earlier, the principle says that the exercise of power “is proper and hence justifiable only when it is exercised in accordance with a constitution the essentials of which all citizens may reasonably be expected to endorse in the light of principles and ideals acceptable to them as reasonable and rational.” (*PL*, p. 217) We also saw earlier that government power is exercised “in accordance with” the democratic constitution of the WOS if those exercising it follow the procedures laid down in the constitution, respect citizens’ rights and liberties, and adhere to the guidelines of public reason. Faced with exercises of power they believe to be unjust, citizens like Jan can still see that those exercises – including the votes and public debates that preceded them -- followed established constitutional procedures and were duly constrained. They can also see that exercises of power like the legalization of abortion or the use of military power were based on reasons of the right kind, even if they themselves would have weighted those considerations differently to arrive at outcomes they regard as more just.<sup>55</sup> They can therefore see that these exercises of power are “in accordance with” the constitution of the WOS and therefore satisfy the liberal principle of legitimacy.

If citizens like Jan recognize legitimacy as a form of political justification, then they can recognize that the laws to which they object are justified. They can do so without having to deny the truth of their religious views and without having to deny that those laws are unjust. This solves the compatibility problem. Rawls thinks once the

---

<sup>55</sup> See *PL*, p. 243 note 32, where Rawls says there is a possible, good-faith weighting of public reasons that would yield a ban on abortion. He does not say that the weighting is reasonable, but he also does not dismiss the questions those who oppose abortion raise about stability for the right reasons.

compatibility problem is solved, it is clear how cases that pose a problem for the stability argument of *PL* are to be handled.

We saw that if justice were the only standard of political justification Jan accepted, then – since the *Ideal of Democratic Governance* requires that exercises of power be justifiable – she might well deny that the enactment of an unjust law realizes the goods of *Democratic Governance*. But if she accepts legitimacy as a standard and recognizes that the law to which she objects still meets it, then the grounds for her denial are removed. Furthermore, the legitimacy of a law is derived from the way that it is enacted. In a democracy like the WOS, a law’s legitimacy is the result of its having been enacted by citizens and legislators following the guidelines of public reason and acting “in accordance with a [democratic] constitution the essentials of which all citizens may reasonably be expected to endorse”.<sup>56</sup> If Jan accepts an unjust law as legitimate, then it must be because she knows that the law was enacted in this way. If she knows that it was enacted in a way that confers legitimacy, then she knows that the exercise of power by which it was enacted was an exercise in *Democratic Governance*. Not only does she lack grounds for denying that the goods of *Democratic Governance* were realized in the difficult case, but she has compelling grounds for affirming that they *were* realized. She can therefore see even the enactment of an unjust law as an instance of “the ideal of citizens governing themselves in ways that each thinks the others might reasonably be expected to accept”.

---

<sup>56</sup> *PL*, p. 217; for the interpolation of ‘democratic’, see *PL*, p. 428.

If realizing that ideal weighs heavily enough with Jan that (3) and (4) are true of her, so that she supports the constitution despite the enactment of the law, then she must attach considerable weight to the legitimacy of the law's enactment. She must not just accept legitimacy as a standard of political justification, she must take it to be an important standard. Once we see what makes an exercise of power legitimate, we can see why she might do that. The on-going legitimate exercise of power under a just democratic constitution realizes the political goods of *Democratic Governance*, and we have seen how great those goods are. If Jan recognizes the importance of these goods and attaches enough weight to them to compensate for the perceived injustice of the law, then (3) and (4) will be true of her and she will affirm her support for the constitution. Rawls sums up his brief discussion of Quakerism in "Reply to Habermas" by remarking that "This [case] illustrates how political values can be overriding in upholding the constitution itself, even if particular reasonable statutes and decisions may be rejected." (*PL*, p. 394) The case of Jan illustrates the summary point that I take Rawls to be making in the remark: that citizens can affirm their support for a constitution while rejecting particular statutes as unjust.

Though Rawls raises difficult cases that are similar to Jan, his discussion of those cases, in the "Reply to Habermas" and elsewhere, is often unsatisfying.<sup>57</sup> Rawls implies, for example, that his treatment of the difficult cases turns on citizens "recogniz[ing] the familiar distinction between accepting as (sufficiently) just and legitimate a constitution ... and accepting as legitimate (even when not just) a particular

---

<sup>57</sup> See also *CP*, pp. 594-95, note 57.

statute or a decision in a particular matter of policy.” (*PL*, p. 393) But he does not elaborate by spelling out exactly why it matters that citizens recognize the distinction to which he refers, nor does he show just how the conclusion about “overriding[ness]” bears on his new treatment of stability. He banks too heavily on the “familiar[ity]” of the distinction and so does not say enough about why citizens might think legitimacy is an important form of justification. As a result, it is hard to see exactly how Rawls thinks the principle of legitimacy helps him cope with difficult cases or what its significance is for the larger project of *Political Liberalism*. By laying out Rawls’s argument for C, by showing how the difficult cases threaten that argument, and by appealing to the *Ideal of Democratic Governance* and its constituent goods, the interpretation provided here suggests some answers.

Of course, on my reading, too, there are still some things about which Rawls should say considerably more.

The most obvious of these can be seen by reflecting on Rawls’s acknowledgement that while legitimacy is less demanding than justice, “[l]aws cannot be too unjust if they are to be legitimate”. (*PL*, p. 428) This suggests that laws cannot be *judged to be* too unjust if they are to be *judged to be* legitimate. If Jan thinks the statute to which she objects is too unjust, she will judge it illegitimate and she may withdraw or qualify her support for the constitution that purportedly legitimated it.<sup>58</sup>

Thus if the notion of legitimacy is to do the work I have said it does, it must still be the case that Jan thinks the statute to which she objects falls within the range of injustice

---

<sup>58</sup> Cf. “[a]t some point, the injustice of the outcomes of a legitimate democratic procedure corrupts its legitimacy.” (*PL*, p. 428)

that legitimacy allows. Rawls seems to acknowledge as much in the summary remark about Quakerism that I quoted earlier, where he contrasts “upholding the constitution itself” with rejecting “particular *reasonable* statutes and decisions”.

But if Jan thinks that a liberal abortion regime or the use of military force is unjust, what would she have to believe about political disagreement on these matters to think that the statutes and decisions to which she objects are still “reasonable” and fall within the acceptable range of injustice? Presumably, she must think that these disagreements are reasonable disagreements, disagreements which are due to the burdens of judgment. (*PL*, p. 55) The problem is that there may be people who deny that disagreements about abortion or war have that status. Since these people deny the burdens of judgment, Rawls could say that they are unreasonable (*PL*, p. 55), that they are therefore not counterexamples to (3) and (4), and hence that their presence in the well-ordered society doesn’t undermine the argument for C. But if he says this, he may seem to be salvaging the stability argument by relying on a response that I said earlier would not do. In that case, introducing the liberal principle of legitimacy merely postponed the point at which Rawls would have to rely on it.

The problem with dismissing Jan as unreasonable stemmed from the possibility that a WOS would include far too many citizens like her. As we saw, that would mean that the argument for C could not show that a justice as fairness and a just constitution would be stable for the right reasons. But even if there are many citizens of the WOS who regard one or another statute as in some way unjust, I think Rawls supposes that the vast majority of them will regard these statutes as legitimate. Once we see what

goods are realized when citizens govern themselves according to a constitution capable of conferring legitimacy, he thinks it plausible that they will regard the legitimacy of the statutes to which they object as weighty enough to override their injustice. There still may be some citizens in the WOS for whom political values are not overriding, but he thinks there would be far fewer than there would be if citizens evaluated statutes and decisions by justice alone. Thus the principle of legitimacy enables Rawls sharply to reduce the number of citizens who count as unreasonable and who merely “go[] along with [the constitution] in view of the balance of political and social forces.” (*PL*, p. xl). If the number is small enough, then despite their presence in the WOS, the argument for C and its consequences shows how justice as fairness can be stable for the right reasons.

This line of thought brings us to another subject about which Rawls should have more to say. Early in *TJ*, Rawls says that, as a theory of generative grammar describes our “sense of grammaticalness”, so “one may regard a theory of justice as describing our sense of justice”. One part of the theory identifies principles which, when applied conscientiously, lead us to everyday judgments of justice. (*TJ*, p. 9) Another part shows how those growing up under just institutions might acquire a sense of justice. One of the most interesting parts of the theory shows how “moral principles can engage our affections” (*TJ*, p. 416), so that citizens of a well-ordered society are moved by their judgments of justice and affirm their sense of justice. This part of the theory depends upon their affections being engaged, according to psychological laws Rawls lays out, by a social ideal: the social union of social unions. (*TJ*, pp. 456-64) According to *TJ*, this engagement of their affections helps to stabilize justice as fairness.

If what I have argued so far is right, then the Rawls of *PL* must have thought that citizens of the WOS would have a sense of legitimacy, as well as – or as part of -- a sense of justice. The Rawls of *PL* identifies a principle which of legitimacy which he says is adopted in the original position, like the principles of justice. That principle, when applied conscientiously, is supposed to lead them to their everyday judgments about the justifiable exercise of power. Moreover, the sense of legitimacy must interact with the sense of justice in complex ways, since unjust exercises of power can be judged legitimate provided they are not judged too unjust. Judgments of legitimacy are supposed to move citizens of the WOS so that they support the constitution and exercise power legitimately themselves when they vote. I believe they are supposed to be moved by judgments of legitimacy because their affections are engaged by a new political ideal: not the social union of social unions, for reasons touched on earlier, but the *Ideal of Democratic Government*. According to *PL*, this engagement of their affections helps to stabilize justice as fairness.

To flesh out this account, Rawls would need to say how the principle of legitimacy is chosen in the original position, how members of the well-ordered society acquire a sense of legitimacy, and what psychological principles explain attachment to the various goods that make up the *Ideal of Democratic Government*. Of course, he offers nothing of the kind. That shows, *contra* the standard reading, just how far *PL* is from providing a theory of legitimacy that is comparable to the theory of justice Rawls developed in his earlier work.

In fact the notion of legitimacy is frustratingly under-theorized. But though it is under-theorized, it enables Rawls to answer what I said earlier is one of the central questions of *PL*:

How is it *possible* for citizens of faith to be wholehearted members of a democratic society when they endorse an institutional structure satisfying a liberal political conception of justice with its own intrinsic political ideals and values[?] (*PL*, p. xl, emphasis added)

Wholehearted membership seemed to be impossible so long as religiously-informed beliefs about justice seemed to be incompatible with the belief that the exercises of power in the difficult cases are justifiable. It is *possible* if justice as fairness is supported by an overlapping consensus of comprehensive doctrines, and if adherents of those doctrines recognize the burdens of judgment, develop a sense of legitimacy, distinguish legitimate from just exercises of political power, distinguish acceptance of the constitution from “accepting as legitimate (even when not just) a particular statute or a decision”, and attach sufficient weight to the *Ideal of Democratic Governance*. Whether that possibility is likely to be realized depends upon details that Rawls does not provide.

### §7. *The Textual Puzzles*

Rather than attempting to fill in details on Rawls’s behalf, I want return to textual puzzles I surveyed at the end of §3 and show that the reading offered here suggests some solutions.

As I noted earlier, a showing of legitimacy is sometimes thought to be a demonstration that states have the right to impose obligations and claim obedience.

This way of thinking about legitimacy is central to the standard reading. Rawls's treatment of political obligation in *TJ* is notoriously controversial.<sup>59</sup> If showing citizens' duty to obey the law were as central to *PL* as the standard reading asserts, we would expect the later Rawls to revisit these topics in his later work. For if the standard reading is right, then the addition of a principle of legitimacy to his theory positioned him to answer the many questions raised about his earlier account. But with the exception of a passage I cited previously in which Rawls says that legitimate laws are "politically (morally) binding" and that citizens are "not to resist [such laws] with force", he seems not to have revisited the subject of political obligation at all.<sup>60</sup>

I believe Rawls touched on the subject only in passing because he did not think he had anything to add to what he had said before. In *TJ*, Rawls had argued that the duty to obey the law follows from the natural duty of justice, for that duty requires us "to support and to comply with" just or reasonably institutions that exist and apply to us. (*TJ*, p. 99) The duty of compliance also "binds us to comply with unjust laws and policies, or at least not to oppose them by illegal means as long as they do not exceed certain limits of injustice." (*TJ*, p. 311)

This last claim and its defense in *TJ* are often overlooked. But in *TJ*, Rawls had said that

---

<sup>59</sup> At *TJ*, p. 98, Rawls says "There is, I believe, no political obligation, strictly speaking, for citizens generally."

<sup>60</sup> *CP*, pp. 578 and 606. Neither the index of *PL* nor that of *Restatement* includes entries for 'obligation' or 'political obligation'; the entries for these topics in the index of *CP* direct readers to papers which antedate *TJ*.

being required to support a just constitution, we must go along with one of its essential principles, that of majority rule. In a state of near justice, then, we normally have a duty to comply with unjust laws in virtue of our duty to support a just constitution. (*TJ*, p. 311)

Thus even in *TJ*, Rawls had clearly stated that citizens of the WOS have a duty to comply with laws they regard as unjust, or not resist them with illegal means, if they satisfy what he would later describe as the conditions of legitimacy. And even in *TJ*, he had clearly implied that if citizens *recognize* the duty of compliance, it is because they *recognize* that laws they regard as unjust satisfy those conditions. *PL* and other late writings return to the duty to obey the law because the difficult cases concern citizens who are subject to laws they might not think themselves duty-bound to obey. If political liberalism is to be possible, it must be possible for these citizens to recognize their duty. To show how they can, Rawls does not provide a new account of political duty and obligation. He simply reminds us of what he said in *TJ*. Citizens in the difficult cases, like other citizens, will recognize their duty to obey an objectionable law if they recognize that the law satisfies certain conditions – conditions to which *PL* draws our attention by labeling them conditions of legitimacy.

If I have correctly identified the really novel work that the principle of legitimacy enables Rawls to do in *PL*, then we can see why he did not dwell on its connection with political duty and obligation. What Rawls says about legitimate exercises of political power helps to show how citizens of the WOS could affirm that their sense of justice and their disposition to support a just constitution belong to their good. But Rawls's treatment of legitimacy was never intended to provide some new argument that citizens of the well-ordered society are obligated to follow directives that

are properly issued by legitimate states, or that legitimate states have claim-rights to the obedience of their citizens.

Another reason Rawls does not connect legitimacy and the claim-rights of legitimate states is, as I indicated in §3, that Rawlsian legitimacy does not attach to states, at least if we understand the state as a society's governing apparatus. It attaches, in the first instance, to constitutions. This is what we would expect if, as I have said, Rawls introduces the principle of legitimacy primarily to help show how it is possible for citizens to affirm the value of their sense of justice and their support for a just constitution, rather than to show anything new about why citizens are obligated or duty-bound to obey a just state.

It is also what the wording of the principle of legitimacy would lead us to expect, at least when the principle is read in conjunction with other of Rawls's remarks. Recall that the principle of legitimacy says, roughly, that political power – such as legislative power -- is exercised legitimately when it is exercised “in accordance with a [just] constitution”. The power to pass laws is “in accordance with” a constitution only if it is exercised “in accordance with” the procedures for passing laws laid out in that constitution, such as majority rule. In “Reply to Habermas”, Rawls predicates ‘legitimate’ of those procedures, saying that “a legitimate procedure gives rise to legitimate laws and policies passed in accordance with it”. Taking the two passages together, we can read Rawls as saying that it is the constitution that is legitimate in the first instance, and that it is a legitimate constitution that “gives rise to legitimate law” when laws are passed “in accordance with it.” More straightforward textual support

comes from Rawls's summary remark about the problem case of Quakerism where, as we saw, Rawls refers to a "legitimate ... constitution" in his treatment of one of the difficult cases.

One text that seems to tell against my reading is the passage in which Rawls says of legitimacy that "reasonable citizens understand this idea to apply to the general structure of authority." (*PL*, p. 393) If we take 'structure' to refer to a part of the part of the basic structure which exercises authority in a WOS, then Rawls would seem to be implying that 'legitimate' applies to the state. But we need not read Rawls this way. Instead, we can read him as saying that it applies to the structure *of* – understood as the *form of* – the state and to the procedures by which governmental authority is exercised. Since these are given by the constitution, we can read the passage as supporting my thesis about the object to which 'legitimacy' attaches.

Finally, I said that one of outstanding questions about legitimacy is what kind of right legitimate states enjoy. One of the puzzling features of *PL* is that it seems to shed so little light on the question of which understanding of legitimacy is the best of those currently on offer. If I am right about Rawls's reasons for making the principle of legitimacy explicit, we can understand why he showed so little interest in that question. For if I am right then, though Rawls's texts allow us to infer that a state exercises a right when it exercises power legitimately, his more pressing interest lies in connecting legitimacy with various axiological notions that are new to *PL* – in particular, with the good of dispositions to respond to the political values that are realized by self-government under a just constitution. If this is right then for Rawls, the preferred

understanding of ‘legitimacy’ prominently includes an element that is quite different than those I canvassed at the outset. Since he takes ‘legitimacy’ to apply, in the first instance, to constitutions and derivatively to laws and directives passed “in accordance with” them, for Rawls, showing that a constitution is legitimate involves showing that it is worthy of support.<sup>61</sup>

### *§8. Conclusion*

I conclude by summarizing the argument. I said at the beginning of §4 that Rawls introduced the conceptual apparatus of *PL* to remedy a difficulty in *TJ*'s treatment of stability, found in Part III, and to answer questions raised by the remedy. In saying that, I meant to deny the standard reading's contention – asserted at its first step - that he introduced that apparatus to shore up an argument for the principles in *TJ*, part I. I also meant to deny one of the radical implications of that reading – namely, that in *PL*, Rawls reoriented his thought around a new concept of which he intended to offer a theory. Rather, I meant to assert that he introduced that apparatus in order to spell out and sustain the stability argument I sketched in §4.

On my reading, a prominent part of the conceptual apparatus of *PL* -- the principle of legitimacy – is deployed to deflect a potential problem with the new stability argument, a problem posed by difficult cases and illustrated by Jan. If that problem can be deflected, and the new stability argument sustained, then a well-ordered

---

<sup>61</sup> Allen Buchanan and Robert O. Keohane impose this condition on legitimacy in their “The Legitimacy of Global Governance Institutions”, a memo prepared for Conference on the Normative and Empirical Evaluation of Global Governance, Princeton University, February 16-18, 2006. The memo can be found at: [http://www.princeton.edu/~pcglobal/conferences/normative/papers/Session1\\_Buchanan\\_Keohane.pdf](http://www.princeton.edu/~pcglobal/conferences/normative/papers/Session1_Buchanan_Keohane.pdf).

society will be stable for the right reasons. Since the well-ordered society which is thereby stabilized is or can be one which complies with the two principles of justice defended in *TJ*, I mean to deny that *PL*'s stress on legitimacy is part of a moral retrenchment on Rawls's part.

The notion of legitimacy deflects difficulties with the new stability argument of *PL* by reminding us of conditions of political duty laid out in *TJ*. Much more interestingly, it helps us to see that just constitutions which enact unjust laws can still realize very great political goods. The most interesting and novel part of Rawls's treatment of legitimacy therefore connects it, not with the right – as on the standard reading -- but with the good. Seeing that helps us gain a more accurate view of Rawls's project in *Political Liberalism* and of the concepts on which the execution of that project depends.