

Legitimacy and Stability in Rawls's Political Liberalism

At the center of his theory of political liberalism, John Rawls placed what he calls "the liberal principle of legitimacy" (LPL):

[P]olitical power is always coercive power backed by the government's use of sanctions, for government alone has the authority to use force in upholding its laws...[Thus LPL says]: our exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason.¹

Rawls's defense of this principle in *Political Liberalism*² has led commentators to ascribe to him the view (1) that legitimacy permits a certain "latitude" in the range of legal arrangements that can be morally binding on a populace;⁴ (2) that the conditions of legitimacy and the conditions of justice "are, at least in principle, independent from one another";⁵ and (3) that what a liberal constitutional democracy "really [has] to worry about is how, when a law is appropriately passed, it is binding on all citizens, even on those citizens who reasonably can differ with it."⁶ Most recently, David Reidy has said that "it would not be misleading to say that 'political liberalism' refers to Rawls's theory of legitimacy," where by "legitimacy" Reidy means a property that a *nation* or *state* exemplifies when it is disposed to enforce laws only when they have been "properly

¹ See Rawls 1996, pp. 136-137; all unadorned parenthetical citations in the text refer to this work.

² (Rawls 1996); all unadorned parenthetical citations in the text refer to this work.

⁴ Wenar 2004, p. 269.

⁵ Brighthouse 1998, p. 721.

⁶ Dreben 2003, pp. 326-7.

enacted or otherwise made law”—i.e. only when they have the right real-world pedigree involving the right sorts of collective decision-making procedures.⁷ A decade prior, David Estlund made a similar claim, declaring that “It is no exaggeration to say that [Rawls’s] two books [viz. *A Theory of Justice* and *Political Liberalism*] are not about the same subject. The first is primarily about justice; the second is primarily about political legitimacy, a topic essentially ignored in [*A Theory of Justice*].”⁸ Finally, Rawls himself appears to have expressed concurrence with these accounts, writing that when citizens’ political power is exercised in accordance with LPL, each should accept “as legitimate (even when not just)” the legislative upshot.⁹

I shall attempt to show that there is strong reason to question these claims about what *Political Liberalism* “is primarily about.” I shall also argue that each of the interpretive statements I have quoted faces a challenge that no commentator has yet met on their behalf. For those who have made these statements each presupposes that “legitimacy” expresses the same property whenever it is used in *Political Liberalism*. In contrast, I claim to identify discussions of legitimacy in *Political Liberalism* and Rawls’s related essays that seem difficult to make sense of if we are understanding “legitimacy” in what I shall call its *enactment* sense, i.e. the sense connected to the idea of a law or policy having been “properly enacted or otherwise made law,” to use Reidy’s explicative phrase. Admittedly, the success of my alternative account would have the unfortunate implication that Rawls himself was led astray by Estlund’s then-unpublished essay when Rawls thanked him for clarifying that legitimacy is primarily a property of democratic procedures and their outcomes, not theoretical conceptions of justice, and then asserted that “legitimacy” is used univocally throughout “this text” (i.e.

⁷ Reidy 2007, p. 248, 246.

⁸ Estlund 1996, p. 68.

⁹ Rawls 1996, p. 393. This statement comes from “Reply to Habermas,” which is included as Lecture IX in Rawls 1996, but which was originally published in 1995, after the first edition of Rawls 1996.

throughout *Political Liberalism*) to express this property.¹⁰ Nevertheless, I will seek to demonstrate that Rawls employs two distinct notions of legitimacy, and that there is no basis for commentators' or Rawls's own claims that whenever legitimacy is discussed in *Political Liberalism*, Rawls is referring to the sort of legitimacy that states exemplify when they enforce only those policies that have the right real-world legislative or administrative pedigrees. In fact, this kind of legislative legitimacy is of only ancillary concern to Rawls in *Political Liberalism*.

Let me begin with a brief statement of what I take to be the interpretive common ground between all commentators (including myself) on Rawls's political liberalism. It is, for example, granted by all that political liberalism is motivated in the first instance by the fact of reasonable pluralism—that is, the fact that informed and conscientious moral reflection will lead reasonable persons to develop starkly different comprehensive moral, religious, and/or philosophical worldviews. Furthermore, it is clearly one of Rawls's aims to argue that the proper response to this fact is a sort of principled restraint on attempts at political justification—citizens should rely only upon ideas that are acceptable to their common human reason, as LPL enjoins. This aim entails that when citizens are guided by “public reason and its principle of legitimacy” (225, 226, 241) in the search for the most reasonable political conception of justice, they must invoke some criterion specifying which principles other citizens “may reasonably be expected to endorse.”¹³

Evidently, this does not mean that citizens may invoke only that which others can be *predicted* to endorse:

[E]ach of us must have, and be ready to explain, a criterion of what

¹⁰ “Reply to Habermas,” in Rawls 1996, p. 429n76.

¹³ Since Rawls never pauses to distinguish *public reason's* principle of legitimacy from LPL, the natural implication is that it just is LPL, which, after all, was presented as *the* (sole) liberal principle of legitimacy.

principles and guidelines we think other citizens...may reasonably be expected to endorse along with us...Of course, we may find that actually others fail to endorse the principles and guidelines our criterion selects. That is to be expected...It is inevitable and often desirable that citizens have different views as to the most appropriate political conception. (226-227)

Since disagreement at this level of thinking about justice is inevitable, we will need a conception of legitimacy in enactment to help us adjudicate the differences that remain even when citizens do their best to heed political liberalism's conditions on responsible political advocacy. Rawls therefore appropriates a familiar democratic procedure to fit within the larger framework of political liberalism: when all citizens sincerely offer what they take to be the most reasonable political conception of justice, "the legal enactment expressing the opinion of the majority is legitimate law."¹⁴

Let me say straightaway that I do not dispute that Rawls uses "legitimacy" here and elsewhere in the enactment sense that is so frequently identified and discussed by commentators. What I do claim, and what I have not seen discussed by others, is that there is strong reason to believe that there is another sense of "legitimacy" at work in *Political Liberalism*. Consider Rawls's remarks describing the relationship between legitimacy and what he terms *stability*. Here Rawls says that we can "make the case that there are adequate reasons for diverse reasonable people" to endorse our specific conception of justice when we can show that "stability...is at least possible" for our conception. "Despite the fact of reasonable pluralism, the conditions for democratic legitimacy are [thereby] fulfilled" (390). "The kind of stability required of justice as fairness is based on its being a liberal view, one that aims at being acceptable to citizens as reasonable and rational, as well as free and equal...Only so is it an account of the

¹⁴ Rawls 1997, p. 137.

legitimacy of political authority..." (143).

These important passages clearly link the ideas of stability and legitimacy. But what does Rawls mean here by "stability"? Most commentators hold that political liberalism's concern with the stability of a conception of justice underlies the political liberal injunction that citizens ought to tailor political justifications so they can win a broad, socially stabilizing consensus among their current reasonable compatriots.¹⁵ On this understanding, there is a clear connection between the ideal of legitimacy in the enactment sense and the goals of social stability and consensus: each is concerned with securing the willing acceptance of those who will be subject here and now to a given law or policy. But Rawls appears to rely on a quite different understanding of stability. He is very careful to stipulate that a conception of justice is *stable* just in case

[1] people who *grow up* under just institutions (as the [candidate] political conception defines them) acquire a normally sufficient sense of justice so that they generally comply with those institutions...[And [2]] the political conception can be the focus of an overlapping consensus...of reasonable doctrines likely to persist and gain adherents *over time* within a *just* basic structure (as the political conception defines it). (141; emphasis added)

And his canonical definition of an *overlapping consensus* is of a consensus that

consists of all the reasonable opposing religious, philosophical, and moral doctrines likely to persist *over generations* and to gain a sizable body of adherents *in...a regime in which the criterion of justice is that [candidate] political conception itself*" (15; emphasis added).¹⁶

¹⁵ Representative examples of this reading can be found in Raz 1990 and Klosko 2004.

¹⁶ At Rawls 1996, p. 10n9, Rawls confirms that this is intended as a definition of *overlapping*

Note that both of these definitions contain what we might call a *developmental qualification*: a conception of justice is stable (in Rawls's technical sense) and is able to win an overlapping consensus when its long-term institutionalization would generate loyalty among those who *grow up* in a society governed by it. If, therefore, stability is a condition of adequacy for a conception of justice, it seems wrong to assess that conception's adequacy by reference to whether its advocacy or implementation would enhance social stability and consensus here and now among citizens who have not grown up under the institutions enjoined by the candidate conception of justice.

The importance to Rawls of testing a conception's capacity to win an overlapping consensus by analyzing the developmental effects of its institutionalization is evident throughout his later works. For instance, consider his response to the worry that the idea of an overlapping consensus on a sufficiently determinate theory of justice is unduly utopian because there are not now "sufficient political, social, or psychological forces" to secure consensus. He suggests that we should understand the question of whether a conception can win an overlapping consensus as the question of how "*over generations the initial acquiescence* in a liberal conception of justice as a *modus vivendi* develops into a stable and enduring overlapping consensus."¹⁷ A *modus vivendi* obtains when citizens are willing to use political power in ways that others reject should they acquire that power, but refrain from doing this out of fear of recriminations once the political tables have turned. So when Rawls says that we can get to an overlapping consensus via the long-term imposition of a stable liberal conception of justice upon reluctant others, this confirms that his main concern is not, as is widely believed, the establishment of political consensus here and now on a specific conception of justice. Rather, he is interested in a

consensus. Other statements of this understanding of an overlapping consensus can be found at Rawls 1996, pp. 10, 141; Rawls 2001b, pp. 390, 410, 414, 421, 430, 479 and 487; and Rawls 2001a, pp. 181, 184, 185, 193 and 194.

¹⁷ See Rawls 2001b, pp. 440-1 (emphasis added). See also Rawls 2001a, pp. 37, 192-3; and Rawls 1996, pp. 158-68.

conception's long-term prospects for "shap[ing]...toward itself," through the political institutions it enjoins, the already reasonable views of those who do not now accept it (389). (Note that Rawls is *not* saying that a given conception of justice has moral authority only for those who have had the right upbringing, or only within those societies that have had the benefit of being shaped by just institutions. His view seems to be that a conception has such authority [for all societies] only if it possesses the *dispositional* property of tending to generate its own support in societies it has the fortune to regulate over generations.)

Granted, this picture will appear mistaken to those who know that Rawls also held that when constructing political justifications, citizens "are to appeal only to presently accepted general beliefs" (224) and "fundamental ideas seen as implicit in the public political culture" (13). Read in light of the stress on "stability" and "consensus", these claims understandably lead many commentators to the view that political liberalism's main aim is consensus and political stability here and now. But my interpretation has the benefit of comporting with Rawls's explicit definitions of *stability* and *overlapping consensus*. And in light of his claim that it was a concern with stability so defined that led him to work out political liberalism in the first place (xvii-xviii, 140ff.), we ought to say either that Rawls misunderstood the implication of his central concern, or perhaps that there are simply two distinct and conflicting concerns at work here.

In fact, this internal tension is evident in Rawls's first sketch of political liberalism, where he claimed both (1) that the ideal of overlapping consensus enjoins us to look for rationales that constitute a "public basis of political agreement", *and* (2) that the question of whether a conception can win an overlapping consensus is "highly speculative" since "we are forced to consider at some point the effects of the social conditions required by a conception of political justice on [citizens'] acceptance of that

conception itself.”¹⁸ Statements along the lines of (1) contribute to the popular view that Rawls tied a conception’s moral authority to its prospects for engendering *present* stability, whereas statements along the lines of (2) support the view that such authority is determined in part through reflection on a conception’s developmental tendencies in certain *hypothetical* circumstances.

Indeed, a strong case can be made that the concern with these developmental tendencies is *the* Rawlsian political liberal concern. After all, it is the developmental notion of stability that is relied upon in Part III of *A Theory of Justice* to test the ultimate adequacy of the theory of justice Rawls developed in Parts I and II, and we are told in the Introduction to *Political Liberalism*, that “all differences” between the account of justice given in *A Theory of Justice* and the account given in *Political Liberalism* result from fixing the problems that the earlier account of justice had meeting this criterion of stability (xvii-xviii). Moreover, in an earlier essay that *Political Liberalism* partly incorporates, Rawls stipulates that a conception of justice is stable just in case it satisfies the developmental criterion, and then asks “Now what more general features of a political conception of justice does *this definition* of stability suggest?” It is only then that he claims—without giving an argument—that one such feature is that “a political conception of justice is formulated so far as possible in terms of certain fundamental intuitive ideas viewed as implicit in the public culture of a democratic society.”¹⁹ It is Rawls’s frequent invocation of this last feature, combined with commentators’ failure to appreciate the special sense in which Rawls is concerned with stability, that has led to the widespread belief that political liberalism’s ultimate goal is enhancing social stability *here and now*. This belief has, I think, prevented commentators from criticizing Rawls’s quick move from the ideal of (developmental) stability to the relevance of ideas implicit in the public culture, despite the fact that the latter seems so out of step with the former,

¹⁸ See Rawls 2001b, pp. 394, 414.

¹⁹ *Ibid.*, pp. 479-81; emphasis added.

which Rawls consistently said was at the heart of political liberalism.

I submit that in light of this understanding of stability, and given the passages, quoted earlier, where Rawls associates (at least one kind of) legitimacy with the ideal of stability, it would be wrong to ignore that there appears to be at least two notions of legitimacy at work in political liberalism. The first is what I have been calling legitimacy in enactment, whereby a law or policy possessing legitimacy in this sense has gone through the legislative or administrative channels required of laws that have a claim to be morally binding on a polity here and now. The second notion of legitimacy is, as I've argued, connected to the stability of a conception of justice, which is a property a conception of justice can possess even if that conception is universally rejected by a populace here and now. Since the conditions of legitimacy in this second sense would seem wholly divorced from conditions that a law or a policy or a conception of justice must meet in order to be morally binding on a populace here and now, this seems to be a different notion than the notion of legitimacy in enactment we began with. The stability-related notion of legitimacy seems better conceived as a notion of legitimacy in advocacy, whereby only conceptions of justice that possess stability are morally licensed for advocacy in a public sphere marked by the conditions of reasonable pluralism.

If this is the right way to understand things, it would entail that it is false to say that *A Theory of Justice* and *Political Liberalism* "are not about the same subject,"²⁰ or that, for Rawls, the conditions of legitimacy and the conditions of justice "are, at least in principle, independent from each other."²¹ Rather, what has emerged as Rawls's stability-centered theory of legitimacy in advocacy will have as much to do with the subject of justice as the condition of stability did in *A Theory of Justice*, where Rawls spent hundreds of pages attempting to show that his theory was fully adequate *qua conception*

²⁰ Estlund 1996, p. 68.

²¹ Brighouse 1998, p. 721.

of justice in part because it was more (developmentally) stable than its competitors.²²

A significant, if somewhat subtle, problem remains for my analysis. Recall Rawls's claim that when we show that a conception of justice is stable, we show that

there are sufficient [public] reasons for *proposing* [that conception, and we thereby demonstrate that] the conditions for [our] legitimately exercising coercive political power over one another...are satisfied...Despite the fact of reasonable pluralism, the conditions for democratic legitimacy are fulfilled. (390; emphasis added; see also 143f.).

Now, *if* Rawls is right that "political power is *always* coercive power backed by the government's use of sanctions" (136; emphasis added), then his statement about the legitimate exercise of political power just quoted seems to commit him to a rather unsavory conception of legitimacy in enactment. For if a collection of theoretical principles of justice passes the developmental stability test; and if passing this test means that the conditions for the legitimate exercise of political power are fulfilled; and if political power is always coercive state power; then legitimacy in advocacy simply collapses into legitimacy in enactment. In that case the normative standards that morally license the *mere proposal or advocacy* of a conception of justice in the public forum will be identical to those that authorize the state's coercive implementation of that conception. If I understand his argument correctly, A. John Simmons rejects political liberalism precisely because he believes Rawls confuses the moral standards associated with ideal justice or responsible advocacy with the standards for legitimate state enforcement: "[T]he Rawlsian argument that shows a type of state to be justified [as an *ideal* of justice] also shows all tokens of that type to be legitimate[ly enforceable at any time]...[L]egitimacy is now grounded not in what [citizens] actually accept or do...but

²² See Rawls 1999a, pp. 347-505.

in what it is reasonable to expect them to accept—that is, in their hypothetical endorsement.”²³ If the passages I just quoted from Rawls tell the whole story, then Simmons’s objection would be devastating to the plausibility of political liberalism.

Of course, I have already quoted passages in which Rawls concedes that when there is actual disagreement within the populace, only “the legal enactment expressing the opinion of the majority is legitimate law.”²⁴ While this may be enough to block Simmons’s charge that Rawls confuses hypothetical endorsement with legitimacy in enactment, further insulation from that charge can be found in textual evidence showing that Rawls in fact rejects (or at least qualifies) his own claim that “political power is *always* coercive power backed by the government’s use of sanctions.” Consider Rawls’s observation that citizens exercise “political power over one another...by voting, if in no other way” (390). Surely if one exercises political power through one’s vote, one does so in only a very qualified sense: one’s vote helps determine coercively enforced public policy *only if* enough others vote the same way. We should therefore distinguish *state political power*, which is *collectively wielded* and *always* backed by the government’s use of coercive sanctions, from what we might call *personal political power*, which is most saliently associated with voting, and which is *never* itself backed by state coercion. This distinction evidently permits the application of distinct moral standards to the evaluation of exercises of the two forms of political power. So while one might be licensed by political liberalism to exercise personal political power on the basis of a (developmentally) stable conception of justice that no one else in fact (currently) accepts, the exercise of state political power should never be permitted on such a basis.

If all of this is on the right track, then each of the two notions of legitimacy that can be found in Rawls’s political liberalism corresponds to a distinct conception of when the use of distinctive forms of political power is legitimate. And yet when Rawls notes

²³ Simmons 1999, p. 759.

²⁴ Rawls 1997, p. 137.

that the enactment sense of legitimacy is the sense in which “legitimacy” is used throughout *Political Liberalism*,²⁵ and when he articulates *the* Liberal Principle of Legitimacy (LPL), there is no indication that Rawls himself sees that both “legitimacy” and “political power” have turned out to have dual meanings in *Political Liberalism*.²⁶

Of course, nothing I have said goes toward establishing the truth of the stability-centered conception of legitimacy in advocacy that I have ascribed to Rawls. In fact, I doubt that such a thing is possible. Still, it is important to get clear about how Rawls’s arguments actually go, especially when his own passing comments about this are sometimes misleading, and sometimes just plain false.³⁰ In light of the importance that so many see in Rawls’s later writings, much more attention should be paid to what is arguably the central controlling concept of *Political Liberalism*, viz. the developmental conception of stability.

REFERENCES

- Brighouse, Harry. 1998. “Civic Education and Liberal Legitimacy,” *Ethics* v. 108, 719-745.
- Dreben, Burton. 2003. “On Rawls and Political Liberalism,” in Freeman (2003), 316-346.
- Estlund, David. 1996. “The Survival of Egalitarian Justice in John Rawls’s Political Liberalism,” *The Journal of Political Philosophy*, 4, 1, 68-78.
- Freeman, Samuel, ed. 2003. *Cambridge Companion to Rawls*, Cambridge, UK: Cambridge University Press.
- Klosko, George. 2004. “An Empirical Approach to Political Liberalism,” in Young 2004, 129-148.
- Rawls, John. 1999a. *A Theory of Justice*, revised edition [first edition: 1971], Cambridge, MA: Harvard University Press.
- Rawls, John 1999b. *Law of Peoples*, with “The Idea of Public Reason Revisited,” Cambridge, MA: Harvard University Press.

²⁵ “Reply to Habermas,” in Rawls 1996, p. 429n76.

²⁶ This suggests a further possibility: that the extension of LPL’s constituent referring phrase, “what others may reasonably be expected to endorse,” also has different interpretations depending on which sorts of political power and legitimacy are at issue.

³⁰ See again both his remarks cited at note 10 and 25, above, and his claim in LPL that “political power is *always* coercive power backed by the government’s use of sanctions.”

- Rawls, John. 1996. *Political Liberalism*, second (paperback) edition [first edition: 1993], New York: Columbia University Press.
- Rawls, John. 1997. "Idea of Public Reason Revisited," reprinted in Rawls 1999b, 129-180.
- Rawls, John. 2001a. *Justice as Fairness: A Restatement*, Erin Kelly, ed., Cambridge, MA: Harvard University Press.
- Rawls, John. 2001b. *Collected Papers*, Samuel Freeman, ed., Cambridge, MA: Harvard University Press.
- Raz, Joseph. 1990: "Facing Diversity: The Case for Epistemic Abstinence," *Philosophy and Public Affairs*, 19, 1, 3-46.
- Reidy, David. 2007: "Reciprocity and Reasonable Disagreement: From Liberal to Democratic Legitimacy," *Philosophical Studies*, 132, 243-291.
- Simmons, A. John. 1999: "Justification and Legitimacy," *Ethics*, 109, 739-771.
- Wenar, Leif. 2004: "The Unity of Rawls's Work," *Journal of Moral Philosophy*, 1, 3, 265-275.
- Young, Shaun P., ed. 2004: *Political Liberalism: Variations on a Theme*, New York, NY: SUNY Press.