

Stability: Political and Conception: A Response to Professor Weithman

George Klosko¹

© Springer Science+Business Media Dordrecht 2015

I am grateful to Professor Weithman for his patient and painstaking response to my criticisms of his book (Weithman 2015). The issues under discussion are of great importance for students of Rawls as well as being inherently interesting. Thus I am pleased to have had this opportunity to think more carefully about them. As I noted in my previous paper, a signal accomplishment of *Why Political Liberalism?* (Weithman 2010) is to lead the reader to see Rawls in a dramatically new light. This is an impressive achievement for an author as much studied as Rawls. Professor Weithman has called attention to a political side of Rawls's political philosophy that has not received adequate appreciation. Although I don't agree with everything Professor Weithman says about the issues in question, I believe he makes a convincing case for Rawls's deep concern with them. However, I believe most readers will view the position Rawls takes on these issues, which Professor Weithman endorses, as extreme and improbable. Central to Professor Weithman's case is the distinction between inherent and imposed stability. Once again, I do not believe the distinction can bear the weight *Why Political Liberalism?* has put on it.

The two main issues on which I focus in this brief paper are the nature of Rawls's concern with the stability of societies and the role stability plays in his justification of justice as fairness.

In responding to my criticisms, Professor Weithman focuses on the distinction between the stability of political regimes and the stability of a conception of justice, and claims that I go wrong in concentrating on the former: 'the stability "Rawls has in mind" attaches to a quite different subject than does the stability Professor

✉ George Klosko
gk@virginia.edu

¹ Department of Politics, University of Virginia, Gibson Hall, 1540 Jefferson Park Avenue, Charlottesville, VA 22904, USA

Klosko says he has in mind' (Weithman 2015, p. 254). 'Rawls was interested in the stability of conceptions of justice and not the stability of "society," or of a "regime" understood as a governing apparatus' (Weithman 2015, p. 254) I believe this claim should be qualified. Although for the most part when Rawls discusses stability he is concerned with conceptions of justice, he is also deeply interested in the stability of societies or institutions. Although I would prefer to avoid becoming embroiled in detailed issues of Rawls-interpretation, evidence for this claim is straightforward. In Section 69 of *A Theory of Justice*, Rawls discusses the stability of systems in general: 'an equilibrium is stable whenever departures from it, caused say by external disturbances, call into play forces within the system that tend to bring it back to this equilibrium state' (Rawls 1971/1999b, p. 457/400). This formula applies not only to conceptions of justice but to social systems as well. In numerous contexts in *TJ* in which stability is discussed, it is undoubtedly in this political sense. For instance, consider the first mention of stability in *Theory of Justice*:

And finally, the scheme of social cooperation must be stable: it must be more or less regularly complied with and its basic rules willingly acted upon; and when infractions occur, stabilizing forces should exist that prevent further violations and tend to restore the arrangement (Rawls 1971/1999b, p. 6/6).¹

For ease of exposition, I will refer to the stability of conceptions of justice as 'conception stability', although a more euphonious term would be desirable, and to the stability of social or political systems as 'political stability.' I believe *Why Political Liberalism* demonstrates that Rawls was not only deeply concerned with political stability but also believed that conception stability could make a significant contribution to it. The view that I have characterized as extreme is Rawls's belief that, in a well-ordered society (WOS), the stability of justice as fairness (conception stability) *itself* would preserve the (political) stability of society. Conception stability accomplishes this, because a stable conception of justice is able to generate its own support. It is 'capable of moving us to compliance for its own sake' (to use Professor Weithman's words), (Weithman 2015, p. 253). We should recognize, however, that when Rawls talks of a conception of justice generating its own support, he is also referring to the contributions of a system of social and political institutions that embody the conception of justice. The contribution of conception stability is especially notable in regard to N-person prisoner's dilemmas. Although coercive institutions are present in the WOS, they generally need not be used. Coercion's main role is to give citizens assurance that other citizens will comply with social norms, thereby allowing them to cooperate, without feeling that they are

¹ Similarly in introducing laws of moral psychology in the original position, Rawls says that, unless a conception of justice is able to generate its own support, 'there would be difficulty in securing the stability of social cooperation' (Rawls 1971/1999b, p. 138/119). See also, e.g., pp. 240/211, 383/336, and 457–458/400–401.

being taken advantage of.² For ease of reference, I will refer to this account of the role of conception stability in contributing to stability in the political sense as Rawls's 'political stability claim' (PSC). Fleshed out a bit, I formulate this as follows:

PSC: Rawls believed that in the well-ordered society, the ability of a conception of justice to generate its own support and so to be complied with for its own sake (i.e., conception stability) is sufficient to generate a spirit of cooperation with political institutions that is able to overcome N-person prisoner's dilemmas (and so to yield political stability).

Fittingly, in view of the two senses of stability, PSC combines two claims. One is psychological, bearing on conception stability. On the basis of the moral psychology discussed in *TJ*, under certain circumstances, a conception of justice will be inherently stable, able to generate its own support. The second claim is sociological, bearing on political stability. Conception stability plays a fundamental role in bringing about general compliance with social norms in collective action situations.³

If we accept PSC as an accurate account of Rawls's views, we should recognize its problems. In spite of Rawls's reluctance to become involved in difficult issues concerning how political and social institutions function in actual societies (e.g., Rawls 1971/1999b, pp. 226–227/199), his claims concerning how conception stability affects people's behavior in collective action situations draws him deeply into this area. In order to see this, it may be helpful to place his views in a context in which they are not often considered. Questions concerning the nature of ideal societies, especially the kind of society in which people can be led to behave morally, have been central to the history of political theory since before the time of Plato. Innumerable theorists who have explored these issues have presented widely different answers. Locating Rawls's views in this context and examining his arguments for his position in comparison to those of a wide range of other theorists leads one to see how utopian his position is—beyond the 'realistic utopia' he claimed to endorse (Rawls 1999a/2001, pp. 11–23/4) I assume it is not necessary to discuss the details of different theorists on these issues, say, Plato, More, Locke, Hume, Bakunin, and Marx, in addition to the 'dark minds of western thought,' Hobbes and St. Augustine, to whom Professor Weithman refers in his book. The extreme nature of Rawls's position seems to me to be obvious. It has been recognized by commentators, e.g., by John Harsanyi. To repeat a quotation used in my previous paper: Rawls's just society 'is not merely an improved version of the best societies now existing; rather it is unlike any society' ever known, because its citizens and

² See Rawls (1971/1999b, pp. 240/211, 267–269/236–237, 576–577/504–505); in ideal theory, general compliance in the WOS is assumed. But in examining stability under realistic conditions, Rawls notes that coercion must also function as a deterrent; see (Rawls 1971/1999b, pp. 575–577/503–505). Still, as noted, even in this context, its main function is providing assurance (Rawls 1971/1999b, p. 577/505).

³ See the passages cited in the previous note: Rawls (1971/1999b, pp. 240/211, 267–269/236–237, 576–577/504–505).

legislators are never motivated by their self-interest 'but rather are always motivated by their strong sense of justice' (Harsanyi 1975, p. 603). It may also be helpful to pause briefly to consider *how* people are led to behave virtuously in different societies. Especially notable is Plato's view that this requires lifelong conditioning and complete domination of society by a philosophical elite who engage in absolute censorship. While Plato's view is unusual in its harshness, we should note how different is Rawls's. According to Rawls, the achievement of justice is without intentional conditioning—beyond the ordinary practices of parenting and socialization in liberal societies. I think it is clear that Rawls's view falls well beyond the bounds of realistic utopia.⁴

Once we understand the sociological side of PSC, we can see how difficult is the challenge Rawls has set for himself. It is not enough to demonstrate the attractiveness of the moral ideal he describes, 'the good of justice.' Moreover, it is not enough for him to demonstrate that the course of moral development he lays out is possible for some people. Rather, he must show that it will be general. All members of society will progress through the three stages—morality of authority, association, and principle (Rawls 1971/1999b, Secs. 70–72). All will acquire desires to be just for its own sake, and so resist temptations of self-interest in collective action situations. There is some evidence that Rawls's position is not quite this strong. In one context, he recognizes that 'there are many who do not find a sense of justice for their good,' and in regard to them, the 'forces making for stability are weaker.' Thus he recognizes an inverse relationship between conception stability and the need for coercion (Rawls 1971/1999b, p. 576/505). However, on even this qualified view Rawls believes that coercive mechanisms 'will seldom be invoked and will comprise but a minor part of the social scheme.' Once again, their 'main purpose' is to give people assurance that their fellow citizens will comply, and so 'to underwrite citizens' trust in one another' (Rawls 1971/1999b, p. 577/505).

Accordingly, there is little doubt that Rawls's position on questions concerning how to bring about moral behavior in society is extreme. Powerful evidence is required to support it. So how strong is Rawls's evidence? He argues according to what he calls 'laws' of moral psychology (Rawls 1971/1999b, Sec. 75). These are empirical claims about psychological development, although, as things stand, they can be neither proved nor disproved by empirical evidence. Because a WOS has not yet existed, Rawls's claims concerning how people would respond to this environment have not been properly tested. In this context, I will not review the process of moral development that Rawls describes. As noted in my original paper, Rawls is quite clear that he is presenting only a sketch of one particular path moral development might take, and that his account draws on 'rather speculative' psychology (Rawls 1971/1999b, p. 461/404). He is aware that a variety of factors can influence development and that people develop in different ways, although he examines only development in a WOS (Rawls 1971/1999b, p. 461–462/404).

Although I do not find Rawls's contention concerning the development of all inhabitants of a WOS persuasive, I will not attempt to provide detailed accounts of possible weaknesses here. The issues are far too wide and complex to be discussed

⁴ On the role of conditioning in bringing about a just society, see Klosko (2003).

adequately in this context. We should recognize that the burden is of course on Rawls to establish that his account is convincing. Whether in fact it is each reader must decide for herself. As noted in my original paper, not only are the issues involved contentious, but if we accept Rawls's burdens of judgment, we should not even hope for agreement on them (Rawls 2005, pp. 54–58). It is enough for me here to establish that (a) Rawls's views on the means to bring about moral behavior in society are unusually optimistic–utopian. In addition, (b) many important questions can be raised about the evidence he provides to support his position. I do not believe there is much doubt about either point.

Even if Rawls's case for PSC has significant weaknesses, it remains to be seen if these have wider implications for our understanding of his theory. This is my second topic. I believe the problems with PSC undermine Rawls's case for justice as fairness. In his response, Professor Weithman correctly notes that the main justificatory argument in *TJ* is based on a pairwise comparison between justice as fairness and average utilitarianism and that, in this argument, stability plays an ancillary role. Rawls's overall argument for justice as fairness is in two stages. In the first, it is compared against other principles of justice in the original position. This is the familiar choice of principles behind the veil of ignorance. Having been chosen, it is tested in the second stage in regard to stability, to see if it will be sufficiently stable to generate its own support. Thus in *TJ*, as Professor Weithman notes, stability is employed as a further test of justice as fairness, after this has been shown to be superior on other grounds (see Rawls 1971/1999b, pp. 6/5–6, 144/124, 504/441, 530–531/465; also Rawls 2001, pp. 88–89, 180) These points I agree with. However, in my response, I could perhaps have been clearer that my criticisms are in regard to Rawls's final and presumably most authoritative justification of his theory in *Justice As Fairness: A Restatement*.⁵ In this work, the role stability plays is much larger. It is a major consideration in the first stage of the argument as well as the second. In other words, it no longer supports an argument based on the advantages of a certain allocation of distributive shares. In *JFR*, considerations of stability themselves bear the bulk of the justificatory burden.

In *TJ*, Rawls had a battery of arguments for justice as fairness, which, he believed, were sufficient to make his case. As just noted, only after the principles were chosen in the first stage, was stability invoked. However, there was a problem. Many readers of *TJ* were not satisfied with his arguments for the difference principle. If the representative individuals in the original principle are motivated to choose the largest possible package of primary goods for the parties they represent, many scholars believe they should choose the package that is on average largest, and so is also larger than that provided by justice as fairness. Rawls was well aware of this problem and responded to it especially with three arguments in Section 26 (Rawls 1971/1999b, pp. 154–155/134).⁶ But many scholars were not convinced. According to Brian Barry, more than any other part of *TJ*, Rawls's defense of the difference principle in particular met with 'uniform rejection' (Barry (1989,

⁵ This is noted, though briefly, Klosko (2015, p. 237).

⁶ I should note that the text of the revised edition in this section is significantly revised.

p. 214).⁷ Rawls's most important argument in Section 26 depends on 'grave risks' one courts in choosing average utilitarianism: this choice has 'outcomes that one can hardly accept' (Rawls 1971/1999b, p. 154/134). However, scholars have noted a simple response. In order to minimize risk, an acceptable minimum could be guaranteed, with economic activity beyond that level unconstrained by the difference principle. Even with a guaranteed floor, average shares would still be larger than those provided by justice as fairness.

In *JFR* Rawls addresses this argument through an additional pairwise comparison, between justice as fairness and what he refers to as 'restricted utility.' This is a mixed conception. Its first principle is the same as in justice as fairness, but in its second principle, it substitutes 'a suitable social minimum' for the difference principle. With a social minimum guaranteed, the 'grave risks' of *TJ* have been eliminated. Without recourse to this line of argument, Rawls's new defense of the difference principle in *JFR* rests much more heavily on stability. He marshals three main grounds for justice as fairness: publicity, reciprocity, and stability (Rawls 2001, Secs. 35–38). It is not necessary to discuss these arguments in detail. The first is mainly concerned with the educative role of a society's principles of justice in the public political culture and how it will help shape inhabitants' conceptions of themselves (Rawls 2001, pp. 120–122). The reciprocity argument is that the difference principle's requirement that those better off not profit from their superior talents unless this also advantages those worst off expresses the most appropriate conception of reciprocity (Rawls 2001, pp. 122–124). The third argument rests on stability. Rawls asserts his now-familiar view concerning the stability of justice as fairness: 'Citizens accept existing institutions as just and usually have no desire either to violate or to renegotiate the terms of social cooperation.' With this established, in accordance with PSC, he moves on to political stability:

Here we suppose that political and social cooperation would quickly break down if everyone, or even many people, always acted self- or group-interestedly in a purely strategic, or game-theoretic fashion. In a democratic regime stable social cooperation rests on the fact that most citizens accept the political order as legitimate, or at any rate as not seriously illegitimate, and hence willingly abide by it (Rawls 2001, p. 125).

The three arguments are intended to demonstrate the superior public culture of the WOS, which is Rawls's main grounds for choosing justice as fairness over restricted utility.⁸ We should bear in mind, however, that these considerations must outweigh the larger average distributive shares provided by restricted utility. Stability therefore is especially important to Rawls, as of the considerations noted in the three arguments,

⁷ According to Barry, in the bibliography of Rawls up to 1981, some three hundred items were on the difference principle and almost all were critical of it (Barry 1989, p. 411 n.1).

⁸ An additional section (Sec. 38) examines problems with restricted utility. I am not able to discuss this here, but it seems clear that one could criticize justice as fairness just as effectively along similar lines. For example, who is to count as the least advantaged, and how do we justify this group over other possibilities?

only appeal to political instability would strongly affect the interests of citizens. If in fact there would be significant differences between the political stability of societies based on justice and fairness and other principles of justice, this would be a significant consideration in favor of the former. Rawls of course contends that the difference would be extremely large, since justice as fairness would generate its own support and so tendencies to injustice would always be overcome. However, in view of the discussion above, it seems imprudent to choose a principle of justice over others that provide larger distributive shares on the basis of so utopian a hope. We should also note that, in the quotation directly above, Rawls appears to assume without argument that societies based on other principles of justice would be unsustainable. Everyone, or at least many people, would 'always' act in self- or group-interested ways in collective action situations. Accordingly, in addition to assuming that inhabitants of the WOS are always just, Rawls apparently assumes that inhabitants of other societies are never just. It would take considerable evidence and argument for Rawls to defend this point, as it is not true that inhabitants of existing liberal democracies always—or even generally—violate collective action norms to pursue their own interests. As noted in my original paper, in the seventy years since WWII all of these societies have been remarkably stable in the political sense. Thus as things stand, in apparently allowing himself this assumption, Rawls makes the argument for justice as fairness much too easy. In the absence of convincing arguments to the contrary, there is little reason to believe that societies based on restricted utility would not be similarly stable. Accordingly, it is not clear that inhabitants' interests are well-served by preferring the smaller distributive shares of justice as fairness over those of other principles, for reasons of stability.

I believe a more accurate account of the role stability should play in the choice of principles of justice requires getting beyond the utopianism of Rawls's position. I think it is beyond question that societies with reasonably just conceptions of justice are able to maintain political stability, although this must be supported by coercive institutions. Because the industrial democracies have institutionalized democratic processes and extensive lists of personal and political rights, they are far removed from the societies imagined by the dark minds as discussed in *Why Political Liberalism?* in which social order is preserved entirely by coercion. In addition to recognizing the conception stability as well as political stability of these societies, we should attempt a more realistic assessment of the stability that would follow from institutions based on justice as fairness. Only if we believe the latter will be markedly higher than the former should this be a significant consideration in the argument from the original position. However, in view of the unquestioned stability of many modern liberal societies, it is unlikely that marginal differences between the stability of the WOS and of them will be a major factor in the representative individuals' deliberations.⁹

⁹ One could advance a related argument concerned with the inherent superiority of a society in which people freely adhere to its principles of justice without the need for coercive institutions, as an expression of their autonomy. But I believe an argument of this kind depends on a particular comprehensive moral and philosophical view and so is ruled out of account. My own feeling, for what it is worth, is that Rawls found it difficult to resist the pull of such an argument and so went too far in trying to substantiate it. As Burton Dreben writes in his excellent discussion of *Political Liberalism*: 'For me, wherever Rawls has gone slightly astray in exposition has been when he has paid too much lip service to Immanuel Kant' (Dreben 2003, p. 340).

And so once again, while I believe that Professor Weithman does a masterful job of explicating aspects of Rawls's political philosophy that had not been understood previously, we should recognize that these portions of Rawls's theory are unrealistic to the point of being utopian. Because they are so implausible, Rawls's views in these areas should not figure heavily in his justification of justice as fairness.

References

- Barry, Brian. 1989. *Theories of justice: A treatise on social justice*, vol. 1. Berkeley: University of California Press.
- Dreben, Burton. 2003. On Rawls and political liberalism. In *The Cambridge companion to Rawls*, ed. Samuel Freeman, 316–346. Cambridge: Cambridge University Press.
- Harsanyi, John. 1975. Can the maximin principle serve as a suitable basis for morality? *American Political Science Review* 69: 594–606.
- Klosko, George. 2003. *Jacobins and utopians: The political theory of fundamental moral reform*. Notre Dame: University of Notre Dame Press.
- Klosko, George. 2015. 'Rawls, Weithman, and the stability of liberal democracy'. *Res Publica* 21: 235–249.
- Rawls, John. 1971. *A theory of justice*. Cambridge, Mass: Harvard University Press.
- Rawls, John. 1999a. *The law of peoples*. Cambridge, Mass.: Harvard University Press.
- Rawls, John. 1999b. *A theory of justice*, Revised ed. Cambridge, Mass: Harvard University Press.
- Rawls, John. 2001. *Justice as fairness: A restatement* Cambridge, Mass.: Harvard University Press.
- Rawls, John. 2005. *Political liberalism*, Expanded ed. New York: Columbia University Press.
- Weithman, Paul. 2010. *Why political liberalism?*. Oxford: Oxford University Press.
- Weithman, Paul. 2015. 'Reply to Professor Klosko'. *Res Publica* 21: 251–264.