

## The Force and Constituency of the Public Justification Requirement: Three Rawlsian Arguments

### INTRODUCTION

Political liberals, aka public reason liberals, accept the “public justification requirement” (PJR). According to the PJR, terms of political cooperation must be acceptable from the point of view of any reasonable citizen. That is to say, they must be such that grounds can be found to accept them from within the framework of any such point of view. Political liberals are of course also liberals. So they think that liberal terms of political cooperation satisfy the PJR, and, indeed, that liberal terms of political cooperation are superior to others precisely because of this. Satisfaction of the PJR isn’t just a welcome bonus for any political order that achieves it: most people who defend the PJR in contemporary political philosophy take it to express or flow from something of fundamental moral importance—the value of justice (Quong 2014), or of respect for others as ends in themselves (Larmore 1999), for example.<sup>1</sup>

The word “reasonable” in the specification of the PJR does idealizing work. It means that satisfaction of the PJR isn’t conditional on actual acceptance among actual people. Political liberals take it to imply certain substantive moral or political commitments—for instance, to recognizing the status of others as individuals with equally good claims to authority in the determination of terms of political cooperation. So the range of points of view from which terms of political cooperation must be acceptable according to the

---

<sup>1</sup> For a survey, see Quong (2013).

PJR (the “justificatory constituency”) does not include the points of view of people who do not share those commitments.

Now, there is a risk associated with building commitments like this into one’s conception of reasonableness. It can look like a kind of gerrymandering of the justificatory constituency of those to whom terms of cooperation must to be justified so as to favor antecedently preferred terms. If you make accepting substantive political principle P a necessary condition of reasonableness, P will turn out to be favored by to the PJR. But of course satisfaction of the PJR may not in that case strike anyone as of much interest. So ideally your conception of reasonableness would incorporate only minimal substantive commitments. In any case it’s important to have a rationale for one’s conception that dispels any suspicion of gerrymandering. At the same time, however, political liberals may need a rationale that justifies incorporating at least somewhat substantive moral commitments into their conceptions of reasonableness, because that may be the only way to ensure that liberal political institutions (uniquely) satisfy the PJR, as they believe them to.

The PJR itself has a Rawlsian pedigree, and Rawls’s own conception of reasonableness builds in fairly extensive substantive commitments. He says that to be reasonable is to propose and be ready to comply with terms of cooperation that secure certain basic liberal freedoms, even at the expense of one’s own interests. He also thinks it’s to be committed (on condition that others are willing to comply too) to fair terms of cooperation, which implies acceptance of principles requiring both equality of opportunity and a social minimum providing for the basic needs of all citizens (Rawls 1996: 48–61, 223–

30; Rawls 2001: 191–2). So Rawls is working with what I’ll call a “restricted-constituency PJR”. He is followed in this by one of the most important contemporary public reason liberals, Jonathan Quong (Quong 2011; Quong 2013; Quong 2014). The intended contrast is with other public reasons liberals such as Gerald Gaus, who start from less restrictive conceptions of the justificatory constituency—including only a shared commitment to the justificatory project alongside conformity to uncontroversial standards of rationality, for example (Gaus 2011: 14–22, 281–2; cf. Gaus 2014: 570).

But, as I’ll explain in this paper, you can reconstruct three distinct arguments for the PJR from Rawls’s writings. And once these arguments are adequately distinguished, a dilemma emerges for Quong and other contemporary proponents of a restricted-constituency PJR who appeal to them. One of the arguments vindicates a restricted conception of the PJR’s justificatory constituency. But this comes at the cost of depriving the PJR of the moral force that contemporary defenders take it to have. On the other hand, the impression of its moral force can be vindicated by either of the other two arguments. But this comes at the cost of expanding the justificatory constituency in a way that threatens restricted-constituency PJR-based defenses of liberalism. The upshot is that contemporary defenders of the restricted-constituency PJR should think twice before invoking Rawls in support of their views. There is no argument to be found in Rawls’s work for any such requirement that also packs the kind of moral punch that contemporary political liberals take it to have.

There is a great deal of contemporary work on the PJR; indeed, something that irritates its critics is the powerful influence it seems to have on contemporary political philosophy (Enoch 2015). And, as I say, Quong is one of the most important writers in the area, and Rawls is invoked by every single contemporary defender of the PJR in any form. Clarity about the three arguments I distinguish sheds light on the PJR and the prospects for PJR-based views. So my argument should be of interest to many, even though its target is only those who accept the Rawlsian restricted-constituency PJR.

Here’s how I proceed. In each of the following three sections (1–3), I distinguish and comment upon one of three arguments for the PJR that can be reconstructed from Rawls’s writings on political liberalism.<sup>2</sup> Then, in the section after that, I set out the dilemma for contemporary defenders of a restricted-constituency PJR. The final section considers two possible ways forward.

#### 1. THE STABILITY ARGUMENT

What I’ll mean by “the stability argument” is an argument for the PJR that can be reconstructed from Rawls’s discussions of the stability of justice as fairness. It’s natural to attribute this argument to Rawls in the context of his presentation of political liberalism as a correction to the argument of *A Theory of Justice*. The basic idea is that justice as fairness, Rawls’s favored conception

---

<sup>2</sup> I distinguish only these three because only these three are fairly explicitly and well worked out in Rawls’s writings. I don’t say that others can’t be found, though I think they’d have to take more enthymematic form.

of liberal justice, must be shown to be justifiable from a wide range of conflicting reasonable points of view if it is to be stable—something that Rawls thinks he failed to show in *A Theory of Justice* (Rawls 1996: xvii–xix).

Before I set out the stability argument, let me say something about the relationship between it and the “liberal principle of legitimacy”. This latter principle (which I’ll discuss more in the sections that follow) offers the best-known formulation of Rawls’s version of the PJR, but it’s not at all prominent in the discussions of stability. One might suppose that it is not conceived by Rawls as part of the conclusion of the stability argument.<sup>3</sup> Nothing I say in this section is intended to imply that it is. Nevertheless, as we’ll see, the PJR can quite naturally be construed that way.

Here, then, is the stability argument.<sup>4</sup> We begin with (1) the fundamental ideas of citizens as free and equal, cooperating together under fair terms, and thence derive (via the original position) the principles of justice as fairness (Rawls 1996: 15–28; Rawls 2001: 5–8, 14–24). We then (2) ask ourselves whether this conception of justice is *viable*. Viability is a matter of fitness to be the public conception of justice for a well-ordered society, which is in turn a matter of (among other things) being stable over time—that is to say, a matter of the society’s institutions remaining just over time by the conception’s standards (Rawls 1996: 15, 35; Rawls 2001: 5, 8–9). Since institutions are maintained over time by the behavior of participants in them, viability is a matter of being

---

<sup>3</sup> Although see Weithman (2010: 312–21; 2015a: 82–6).

<sup>4</sup> In broad outline, my account here follows that of Weithman (2010: chapters 9–10).

capable of eliciting just-institution maintaining behavior (of the conception’s “generating its own support”). Next, we note that (3) one way to elicit the requisite behavior employs oppressive means—means that violate the basic liberal freedoms of at least some people. And we further note that (4) if viability can be secured only because liberal institutions are capable of eliciting the requisite behavior through oppressive means, then (paradoxically) liberal institutions are viable only insofar as they are illiberal (Rawls 1996: 36–8, 142–3; Rawls 2001: 33–4, 185). So we arrive at the interim conclusion that (5) justice as fairness is viable only insofar as institutions satisfying its standards are capable of eliciting just-institution maintaining behavior, but without employing oppressive means to do so.

Now Rawls goes on to argue (6) that the institutions of justice as fairness *are* capable of non-oppressive elicitation of the requisite behavior, adducing in support of this conclusion (a) an account of moral psychology under liberal institutions (Rawls 1996: 81–6, 141, 163; Rawls 2001: 184, 195–7), (b) the importance of the “characteristic values of the political” as expressed by the principles of justice as fairness (Rawls 1996: 156–7, 160–2; Rawls 2001: 189–90), and (c) a description of a sufficiently wide and realistically attainable “overlapping consensus” on the “political conception” of justice as fairness (Rawls 1996: 164–72; Rawls 2001: 190–5).<sup>5</sup>

---

<sup>5</sup> Arguably, Rawls also adduces (d) the fact that citizens will regard the liberal principle of legitimacy as an important value and will be assured that the well-ordered society of justice as fairness (in which not only the principles of

But we can ignore (a)–(c) for the moment. The stability argument for the PJR is complete in its essentials by (5), given two assumptions that Rawls makes. The first we can call

“the necessity of endorsement”: that “an enduring and secure democratic regime...must be willingly and freely supported by at least a substantial majority of its politically active citizens” (Rawls 1996: 38; Rawls 2001: 34).

The second and more familiar is

“the fact of reasonable pluralism”: that “[u]nder the political and social conditions secured by the basic rights and liberties of free institutions, a diversity of conflicting and irreconcilable yet reasonable comprehensive doctrines will come about and persist” (Rawls 2001: 34).<sup>6</sup>

---

justice but also the values of public reason are adhered to) satisfies it. See Weithman (2010: 327ff), and below.

<sup>6</sup> See also Rawls (1996: 36). A comprehensive doctrine is a system of ideals, principles, and standards, worked out to cover “what is of value in human life, and ideals of personal character, as well as ideals of friendship and of familial and associative relationships, and much else that is to inform our conduct, and in the limit...our life as a whole” (Rawls 1996: 13).

Given these two assumptions, (5) implies two things. First, that a sufficient number of citizens must endorse the institutions of justice as fairness simply because those institutions seem justified to them. Second, that the necessary support will not be forthcoming unless the institutions can be given justifications from within as wide a range of comprehensive views as are represented among a group of citizens that is large enough for their endorsement to be enough to sustain institutions meeting the standards of justice as fairness.<sup>7</sup>

Steps (1)–(5) of the stability argument, together with the necessity of endorsement and the fact of reasonable pluralism,<sup>8</sup> imply that there is *no* chance that justice as fairness (or, indeed, any liberal conception) will be viable if it depends for its plausibility on the truth of any one comprehensive doctrine in particular. There just won’t be sufficient numbers of people who accept that comprehensive doctrine for the requisite degree of institution-supporting behavior to be elicited. To affirm it anyway would be unrealistically utopian,

---

<sup>7</sup> More precisely: since under liberal institutions, people’s comprehensive views may be likely to evolve over time, what’s needed is justifications from within a sufficiently wide range of comprehensive conceptions that at any one time it is likely that a sufficiently large proportion of the citizenry will adhere to conceptions from within this range to sustain institutions meeting the standards of justice as fairness.

<sup>8</sup> On a natural interpretation according to which it won’t be the case that a single comprehensive doctrine stably persists even only among the majority.

since it would require us to disregard either the fact of reasonable pluralism or the necessity of endorsement.

So what's needed, if justice as fairness is to be viable, is that the principles of justice as fairness at least cohere with and ideally find support within each of a sufficiently large number of the diverse range of views that can be expected to persist under institutions meeting its standards.

As I said above, Rawls adduces considerations (a)–(c) in support of the viability of justice as fairness. That is, he explains how it can be affirmed from within a wide range (the overlapping consensus) of reasonable, realistic comprehensive views, and he suggests that a combination of psychological forces and the significance of the political values it expresses (which must be interpreted here as their significance *from the point of view of citizens*)<sup>9</sup> will tend to encourage and sustain that affirmation and an accompanying endorsement of the institutions of justice as fairness.<sup>10</sup> Effectively, then, he suggests that we can realistically anticipate that a sufficiently large proportion of citizens will maintain comprehensive views within the overlapping consensus for their endorsement to secure its stability. (This might be doubted, but it's not important for my purposes to address the doubts.) And we can't be sure enough that any proper subset of the views in question will be held by a sufficiently large majority that the others can be safely disregarded.

---

<sup>9</sup> This isn't always clear from the way that Rawls presents things at e.g. Rawls (1996: 156–7) or Rawls (2001: 189–90). Compare Weithman (2015b).

<sup>10</sup> But, as Weithman (2015b) argues, he ends up worrying about this argument. See also below.

We can't, for example, appeal to the fact that the justice as fairness coheres with utilitarian and Kantian views and stop there, since utilitarians and Kantians together can't realistically be expected to be in a sufficiently large majority.

From this we can infer a version of the PJR. Justice as fairness is viable only if it can be justified from within as wide a range of comprehensive views as are represented among a group of citizens that can be expected to be large enough for their endorsement to be sufficient to sustain institutions meeting the standards of justice as fairness. The group of *reasonable* citizens—that is, citizens who affirm the political values that justice as fairness expresses—can be expected to be large enough for their endorsement alone to be sufficient. And no other group (including proper subsets of the group of reasonable citizens as well as groups of reasonable and unreasonable citizens) can be expected both to endorse the standards of justice as fairness *and* be large enough their endorsement alone to be sufficient. So justice as fairness is viable only if it can be justified in terms that everyone within the restricted constituency of the overlapping consensus on the political values can affirm. For only then are its institutions likely enough to generate the requisite support, the principles of justice as fairness being such as to gain the endorsement of a sufficiently large proportion of the citizenry, despite the fact of reasonable pluralism. Hence the PJR.<sup>11</sup>

---

<sup>11</sup> I've left out a great deal of detail here—for example, the detail that would be needed to explain why the justificatory constituency must, according to the stability argument, converge on the political values expressed by justice as

*Comment*

The stability argument for the PJR makes sense of a restriction of its justificatory constituency, because we're concerned with the society of justice as fairness and the citizens thereof, and in that society stability needs the endorsement only of those who affirm the political values. (Perhaps we can also generalize to liberal societies more generally.) But it's a rather unsatisfying argument. The PJR looks like a fairly fundamental moral-political principle, and Rawls and Rawlsians often treat it that way. In particular, it looks like a principle whose violation constitutes a grave wrongdoing—of those to whom the terms of cooperation are thereby shown not to be justified, or of those who do worst under the terms of political cooperation that are thereby shown not to be justified, for example. But it's hard to see how the concern for stability could yield this kind of conclusion. Who is wronged if justice as fairness, though unviable in the sense that's in play for the stability argument, is implemented?

It helps a bit to remember how much in the stability argument for the PJR depends on the assumption that the principles of justice have already been established as presumptively sound, that is, as appropriate given the fundamental ideas of freedom, equality, and fairness. It's because they're taken to have been established that the mechanisms of stability must be non-oppressive. So the PJR is derived in part from a prior commitment to liberal

---

fairness rather than on the principles of justice as fairness alone. I've left it out because it's not necessary to make the points I want to make.

justice, and in particular to the basic liberties, given that these are what oppressive means of securing stability violate.<sup>12</sup>

Nevertheless, given the work that empirical assumptions do in completing the stability argument for the PJR, you might still be unsatisfied. There's the assumption of the necessity of endorsement, the assumption that reasonable pluralism precludes dependence for stability on anything deeper than the political values, assumptions about moral psychology, and assumptions about demographics. So if justice as fairness (or more generally liberal justice) turns out not to be viable because it doesn't satisfy the PJR, the stability argument implies that the problem is one of the sociology of liberal institutions, not that liberal institutions violate some independently grounded right. Granted, the sociology gets its significance from the fact that political theorizing must "fall under the art of the possible" (Rawls 2001: 185). But this still doesn't seem to pack the moral punch that you'd naturally think accompanied the PJR—the sort of punch that makes it clear that those to whom the principles can't be justified are wronged by their imposition.<sup>13</sup>

---

<sup>12</sup> This is the interpretation of "stability for the right reasons" that's most natural in the context of Rawls's main discussions of that notion at (Rawls 1996: 142–4; Rawls 2001: 185–6). An interpretation that takes it to presuppose the PJR is not.

<sup>13</sup> Compare G.A. Cohen's doubts about constructivism (Cohen 2008).

## 2. THE DEMOCRATIC CITIZENSHIP ARGUMENT

What I'll call "the democratic citizenship argument" is the one that most resembles contemporary proposals about the ground of the PJR. It is presented in *Political Liberalism* and *Justice as Fairness: A Restatement* as a foundational motivation for political liberalism, although in neither book does it occupy center stage in the development of the narrative.

The argument is relatively simple. First, we note some facts about the distinctive nature of the political relationship. It's effectively involuntary and it's coercive (Rawls 1996: 135–6; Rawls 2001: 40, 93–4). At the same time, political power is regarded as the power of free and equal citizens as a collective body.<sup>14</sup> And yet many citizens may not regard much of what the state does as justified. These facts are said to give rise to a puzzle about political legitimacy, which is made all the more pressing by the fact of reasonable pluralism. The puzzle is to explain "in the light of what reasons and values—of what kind of conception of justice—citizens [can] legitimately exercise that coercive power over others" (Rawls 2001: 41). This needs explaining because a threat to citizenly equality looms: if no suitable justification can be found, then the exercise of political power will amount to nothing more than the domination of some by others.<sup>15</sup> And a justification that simply appeals to the truth won't do, because to appeal to a truth that freely reasoning others don't accept is to fail to acknowledge the "burdens of judgment" that make epistemically faultless

---

<sup>14</sup> At least in a constitutional democratic regime (Rawls 1996: 136; Rawls 2001: 40).

<sup>15</sup> Compare Rawls (1999: 578).

disagreement about value the inevitable outcome of the free exercise of human reason (Rawls 1996: 60–1).<sup>16</sup>

Although it's not clear from some of what Rawls says,<sup>17</sup> the fact of reasonable pluralism isn't necessary to generate the puzzle of political legitimacy. The puzzle arises even if everyone shares a single comprehensive view, so long as there is sometimes (reasonable) disagreement about which state actions are justified according to this view—as there might be, since even people who agree on values and principles may disagree on the way in which these apply in practice.

The puzzle is solved, Rawls suggests, if and only if legislative power to enact the disputed policies is exercised in accordance with a constitution (a "general structure of authority") that everyone can endorse, even though they can't all endorse the policies (Rawls 1996: 136–7; Rawls 2001: 40). If everyone shares a single comprehensive view, the solution involves finding sufficient reasons within the shared view for endorsing the constitution. Of course, the fact of reasonable pluralism implies that there is no shared comprehensive view, but that doesn't change the essential structure of the proposed solution. It just makes it look more empirically unlikely, for it means we can't look to any shared comprehensive view for the reasons to endorse the constitution. We have to restrict ourselves to what "all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their

---

<sup>16</sup> For Rawls's account of the burdens of judgment, see Rawls (1996: 54–8; 2001: 35–6).

<sup>17</sup> For example, at Rawls (2001: 40).

common human reason” (Rawls 1996: 137), and this calls for the exclusion of comprehensive values altogether. Either way, we get some version of the PJR. Satisfying it is necessary to reconcile the idea that terms of political cooperation are collectively self-imposed with the distinctive kinds of disagreement that are inevitable under liberal institutions, as the ideal of democratic citizenship requires.

*Comment*

This argument highlights a tension between the facts of involuntariness and coercion<sup>18</sup> (henceforth, for brevity, I’ll just talk of involuntariness) and the idea of collective self-imposition, or more precisely the idea that no one is to be treated as subordinate to anyone else.<sup>19</sup> But it prompts doubts insofar as it is taken to justify a restricted-constituency PJR. As I said above, Rawls takes members of the relevant justificatory constituency to affirm basic liberal freedoms and both equality of opportunity and a social minimum providing for the basic needs of all citizens (Rawls 1996: 227–30). In short, they accept the liberal political values of freedom, equality, and fairness.

The problem is that the democratic citizenship argument doesn’t seem to justify any such restriction of the justificatory constituency. Why must we look only for justifications that will satisfy people who are reasonable in this sense? That is: why would we think that justifications in terms of the political values

---

<sup>18</sup> We might add the fact of the profoundness of the basic structure’s effects. See Rawls (2001: 55).

<sup>19</sup> Note removed for blind review.

are sufficient to solve the puzzle of legitimacy?<sup>20</sup> Even in a liberal democracy—perhaps *especially* in a liberal democracy—there will be people who don’t accept these values. And they’re involuntary participants in the political scheme too. How can the idea of collective self-imposition be reconciled with *their* involuntary participation? So the democratic citizenship argument creates a pressure to widen the justificatory constituency of the PJR beyond that specified by Rawls.

A defender of the democratic citizenship argument might respond to this concern by saying that her interest is in ideal theory, and in particular that she is concerned only with the liberal society under idealized conditions. Such conditions, she might continue, include universal acceptance of the liberal political values. In that context, resolving the tension between collective self-imposition and citizens’ involuntary participation calls for satisfaction only of a restricted-constituency PJR. The case of people who don’t accept the liberal political values is not relevant.

Now, it’s reasonable to ask here what the rationale is for the interest in ideal theory, and what kind of idealization the rationale supports. One possibility is that we are interested in discovering how a liberal conception of justice fares under favorable but realistic conditions. This would explain why we should not idealize away the fact of reasonable pluralism: to do so would be unrealistically utopian. So we have to resolve the tension between the idea of collective self-imposition and the involuntary participation of those who share the liberal political values but differ in their wider comprehensive views. But we don’t

---

<sup>20</sup> As Rawls clearly thinks they are: see e.g. Rawls (2001: 91).



have to worry about resolving the tension between the idea of collective self-imposition and the involuntary participation of those who reject the liberal political values, because these people can be assumed away.

The problem with this way of interpreting the democratic citizenship argument is that the proposed rationale doesn't justify assuming away people who reject the liberal political values. For such people are also inevitable under favorable but realistic conditions (Rawls 1996: 39, 65). To assume them away would be unrealistically utopian. A different rationale for idealization will need to be found if it's to justify the restricted-constituency PJR.

So instead you might argue that the ideal of democratic citizenship and the associated idea of collective self-imposition don't apply to the relations between any two citizens, regardless of their attitudes to one another. Appealing to ideas of reciprocity, you might suggest that a person effectively renounces her right to participation in collective self-rule insofar as she fails to accept the reasoning that gives rise to the PJR and with it the need for terms of cooperation to be justified to everyone else. A person is not reasonable if she is not willing to compromise. So we don't need to expand the justificatory constituency so far as to include even those who do not affirm others' status as having equal authority in the determination of terms of cooperation despite their differing comprehensive views. Such people can be idealized away (Rawls 1996: 48-50; Gaus 2011: 14-22).

This is plausible. But it doesn't offer a way to resist the expansionary pressure of the ideal of democratic citizenship altogether. It's hard to see how that ideal can be plausibly supposed to have application only to the relations

between individuals who accept the liberal political values in particular. The restriction of the justificatory constituency warranted by the reciprocity-based approach isn't extensive enough to exclude everyone who disputes the kind of liberalism that proponents of a restricted-constituency PJR take to be justified. Recognizing that our terms of cooperation must be justifiable in terms that each who recognizes the same thing can affirm, given their views, does not entail affirming the liberal political values. That depends on the content of the views of the people who recognize the need for such justification.<sup>21</sup>

### 3. THE ORIGINAL POSITION ARGUMENT

At one point, Rawls suggests that the democratic citizenship argument can be restated "more rigorously" by looking at the question of legitimacy from the point of view of the original position. To do so, he says, "[w]e suppose the parties to know the facts of reasonable pluralism and of oppression along with other relevant general information. We then try to show that the principles of justice they would adopt would in effect incorporate this principle of legitimacy and would justify only institutions it would count legitimate." (Rawls 1996: 137; cf. Rawls 1996: 62)

The proposal is carried out in Rawls's discussion of the idea of public reason. There, he explains that the parties in the original position "would not act

---

<sup>21</sup> Compare Gaus (2011: 276-83). There is another way to reconceive the democratic citizenship argument so that it invokes an ideal that applies only within the context of the relationship between people who already accept the liberal political values. I discuss this reconception below.

responsibly as trustees” if, “[i]n securing the interests of the persons they represent”, the parties did not “insist that the application of substantive principles be guided by judgment and inferences, reasons and evidence that the persons they represent can reasonably be expected to endorse ... Thus we have the principle of legitimacy.” (Rawls 1996: 225)

Why would the parties not act responsibly as trustees? Rawls simply says that “When equally represented, no citizen could grant to another person or association [the] political authority [to decide constitutional essentials as that person’s, or that association’s, comprehensive doctrine directs]. Any such authority is, therefore, without grounds in public reason, and reasonable comprehensive doctrines recognize this.” (Rawls 1996: 226; cf. Rawls 2001: 191–2) This is the original position argument.

#### *Comment*

This argument is puzzling. The parties in the original position know that those they represent have an interest in living according to their freely chosen conceptions of the good, which is what dictates the choice of the first principle. So clearly they couldn’t grant anyone the authority to decide on constitutional essentials that are incompatible with the first principle. Indeed, they presumably couldn’t grant anyone the authority to decide on constitutional essentials that are incompatible with either of the two principles, assuming that their reasoning for those principles succeeds.

Given that it may not always be clear how and whether the principles of justice apply in any given case,<sup>22</sup> this line of thinking may well give the parties stability-based reasons to affirm the “values of public reason”:

not only the appropriate use of the fundamental concepts of judgment, inference, and evidence, but also the virtues of reasonableness and fair-mindedness as shown in the adherence to the criteria and procedures of commonsense knowledge and to the methods and conclusions of science when not controversial.” (Rawls 2001: 92)

The idea would be that applications of the principles of justice must not seem to depend on controversial views if there is to be the widespread endorsement of the institutions supposedly implementing them that is necessary for the long-term stability of social justice.

But the values of public reason are designed to ensure that people can see that the principles of justice are satisfied by concrete institutions. The legitimacy of the principles themselves is assumed. So the point that the parties aim to protect the interests of those they represent doesn’t justify the PJR, which is understood by contemporary public reason liberals to specify a condition to be met by the principles themselves, not merely ways of implementing them.<sup>23</sup>

---

<sup>22</sup> See Rawls (1996: 223–4); compare Ripstein (2009: 170).

<sup>23</sup> Although see Weithman (2010: 319).

Of course, those whom the parties represent are taken to have an interest in living according to their freely chosen conceptions of the good. So that might be thought to justify the PJR. The parties might reason that, although the principles of justice seem to them to be the best way of advancing the interests of those they represent, if there are citizens to whom the principles cannot be justified in terms that they accept, then those citizens' interests won't have been served after all. (They'll be subjected to an effectively inescapable framework that reflects ideals that they cannot reconcile with their own.) Whatever the merits of this line of reasoning, however, it will give rise to the pressure to expand the justificatory constituency beyond its Rawlsian specification once again. If the concern is real, it's a reason to worry about acceptability to citizens who lack the substantive commitments of Rawlsian reasonableness as well as those who have them.

This problem would go away, of course, if it could be stipulated that the parties take those they represent to accept the political values. And the parties *do* know that they have a capacity for the sense of justice. But whether that capacity will be realized in an acceptance of the principles of justice and the values that they express is part of what's supposed to be determined by the stability argument. And even that argument doesn't purport to show that *all* citizens will accept the principles of justice. (Just a sufficiently large number.) So the parties' knowledge of the capacity for a sense of justice can't be the basis for their selection of a restricted-constituency PJR.<sup>24</sup>

---

<sup>24</sup> The suspicion that I'm raising here that the original position argument by itself can't justify the PJR is deepened by the fact that Rawls's presentation of

#### 4. A DILEMMA FOR DEFENDERS OF THE RESTRICTED-CONSTITUENCY PJR

Now that we've reviewed the three arguments, we can see that Rawls-inspired defenders of a restricted-constituency PJR face a dilemma. The original position argument and the democratic citizenship argument don't look capable of justifying the restriction of the relevant justificatory constituency. The stability argument avoids this problem, but it fails to provide the PJR with the moral force that its proponents take it to have.

There's a temptation to avoid the dilemma by effectively offering one of the arguments for the PJR while importing the restriction of the justificatory constituency from another. The most promising-looking way of doing this would be to use the stability argument to justify the restriction but one of the other two to provide the moral force that's otherwise lacking. But this won't work, because the moral force that either of the other two arguments furnishes for the PJR can't be restricted in the right kind of way by the considerations invoked by the stability argument. As we saw, the stability argument explains the restriction of the justificatory constituency as follows. Justice as fairness must be stable, and it won't be stable if it's not justifiable in terms of the political values. It doesn't need to be justifiable to anyone who doesn't accept those values, because the overlapping consensus on those values will be sufficiently populous for stability without the endorsement of anyone else. There's *no moral* reason here for thinking that the endorsement of others isn't the argument in places seems to presuppose the principle of legitimacy. See Rawls (1996: 224; 2001: 89).

relevant; it's just that we can do without it for stability. Yet what's needed is precisely a moral reason—something that can defeat the prima facie implication of the democratic citizenship and original position arguments that citizens outside the restricted constituency are owed justification as much as those within it are.

I think Jonathan Quong may succumb to the temptation I'm describing here. He defends the restriction of the relevant justificatory constituency by arguing for what he calls an "internal conception" of political liberalism. According to the internal conception, the justificatory problem that reasonable pluralism poses is a problem only for a well-ordered Rawlsian society, in which everyone accepts the political values (Quong 2011: 143–5; cf. Rawls 1996: 35 and Rawls 2001: 8–9). So anyone who doesn't is excluded from the justificatory constituency from the outset. The idea is that we can now run one of the arguments with more moral force and yet help ourselves to the justificatory restriction.

But why are we interested in the well-ordered society? Here's what Quong says:

If liberalism cannot gain the adherence of citizens under ideal conditions, it would be unstable and self-defeating. Justifying our liberal principles to reasonable citizens [i.e. those who accept the political values—TS] is thus...an effort to probe "the limits of practical possibility", and thus show that a stable, well-ordered liberal society is not an impossible dream, but is rather "realistically utopian". (Quong 2011: 160)

According to this, the restriction of the PJR's justificatory constituency is motivated by the stability argument. Yet that sits uneasily with Quong's official position that the ground of public reason is justice, which suggests rather more moral force than the stability argument provides:

Ensuring that the principles that regulate our shared political life can be justified to each citizen by reference to this [publicly recognized] common view [of society as a fair system of social cooperation between free and equal citizens]—showing these principles can be the subject of public reason—is how we can know that the terms that regulate our political institutions are fair, and therefore just. (Quong 2013)<sup>25</sup>

Now, strictly speaking, what Quong says here isn't in tension with a stability-based argument. Since the stability-justified PJR requires that principles of political cooperation be justifiable in terms that all political values-accepting citizens can accept, and since liberal institutions, which are assumed to be just based on familiar arguments from freedom and equality, uniquely satisfy that condition, it follows that ensuring satisfaction of the stability-justified PJR is a way of knowing that the terms that regulate our political institutions are just.

---

<sup>25</sup> See also Quong (2014: 273–5). I read Quong (2011) as endorsing a justice-based conception of public reason, but his defence of the view there is not explicitly set out.

But their satisfaction of the PJR isn't the *reason* that they're just, which is what the claim that justice is the ground of public reason would suggest.

Just to be clear: there's nothing amiss about asking, as Quong effectively does, whether an idealized liberal society will be one in which the ideal of democratic citizenship is upheld, in which no one is dominated by anyone else. And if such a society is one in which everyone can be presumed to accept the political values, then our question will take the form of asking whether the restricted-constituency PJR is satisfied. The point is that it's hard to arrive at the PJR in this way without supposing that satisfaction of the ideal of democratic citizenship matters generally—that non-domination is in general a fundamental political virtue.<sup>26</sup> But then the constituency expansion problem lurks in the background, even if it's not going to trouble you in the context of your particular interest in the idealized liberal society. You're committed to justifying even liberal political arrangements in general to people who don't accept liberalism, so long as such people exist under liberal institutions—as we can assume they will.

It's easy to miss the force of this worry in reading Quong's work.<sup>27</sup> He focuses on the well-ordered liberal society in particular because he's interested

---

<sup>26</sup> It's not impossible, though, as the possibility of the alternative construal of the democratic citizenship argument shows. See section 2 above and section 5 below.

<sup>27</sup> And Rawls's. Neither explains as clearly as he might that a well-ordered society in which everyone accepts the political values is only the *starting point* in our stability thought experiment—from which things can be expected to

in how liberalism fares under idealized conditions. But one could be interested in how it fares in either of two ways. One could be interested in how it fares from the point of view of stability. And one could be interested in how it fares from the point of view of the ideal of democratic citizenship. Both interests would make it rational to ask whether the restricted-constituency PJR is satisfied. Only the latter gives us reason to think that the PJR embodies an ideal with the kind of moral power that would be implied by a basis in the value of justice. But it tends to commit us at the same time to an unrestricted account, one that Quong explicitly repudiates. If Quong's interest in public reason derives from only an interest in how liberalism fares under idealized conditions from the point of view of stability, as *Liberalism without Perfectionism* suggests, then he's entitled to the repudiation, but it looks as if he should give up the suggestion that public reason is grounded in anything so morally important as justice.

##### 5. THE PROSPECTS FOR THE RESTRICTED-CONSTITUENCY PJR

What options are open to Rawlsian defenders of the restricted-constituency PJR? I conclude by considering two, both of which begin with the ideal of democratic citizenship.

The first reconceives the ideal of democratic citizenship as an ideal that applies (at least insofar as it generates the PJR) only between individuals who accept the liberal political values in particular, because such people stand in a

---

deteriorate somewhat in the sense that *some* unreasonableness is sure to develop. The question is whether *too much* will develop.

distinctive kind of relationship to which that ideal is internal.<sup>28</sup> There is no need, it says, to be concerned with the tension between the idea of collective self-imposition and the involuntary participation of citizens in the case of citizens who reject the liberal political values, because appealing to the ideal in such a way as to generate the tension makes no sense in such cases. And so the democratic citizenship argument creates no pressure to expand the constituency of the PJR after all.

By analogy, think of the way in which friends must comply with certain rules of conduct in order to live up to the ideal of friendship in their interactions. Since that ideal doesn't apply to interactions between non-friends, it would be strange to infer that the rule did. Suppose that there is some friendship-specific analogue of the PJR. An idealization that restricted its justificatory constituency to people who accepted the values of friendship would surely not be inappropriate.

However, this approach tends to undermine the impression that the satisfaction of PJR (as justified by the democratic citizenship argument) constitutes a fundamental political virtue. To see this, consider the following mistaken objection. The objection is that, on this interpretation, the democratic citizenship argument makes satisfying the PJR irrelevant for the purposes of assessing liberal institutions in non-ideal cases. Many actual citizens in liberal democracies will not accept the liberal political values. These citizens won't have the relationship with one another that makes the democratic citizenship ideal applicable, and so the democratic citizenship

---

<sup>28</sup> Note removed for blind review.

argument on as we're now conceiving it gives us no reason to worry about reconciling the idea of collective self-imposition with their involuntary participation.

Here's why the objection is mistaken. Suppose that only two actual citizens accept the liberal political values. The ideal of democratic citizenship does (by hypothesis) apply to *their* relations, and so satisfaction of the PJR is needed to show that the tension between the idea of collective self-imposition and *their* involuntary participation is resolved. So it is after all relevant for assessing liberal institutions in (most) non-ideal cases.

But this serves to highlight the way in which the idealizing strategy of the envisaged interpretation of the democratic citizenship argument appears to limit its applicability. For the ideal of democratic citizenship doesn't seem to do any work now in justifying the imposition of the liberal principles with respect to anyone *other* than our two citizens. The imposition of liberal principles generally must be justified in some other way; the argument here tells us little more than that if they are imposed, they will make it possible for citizens to live up to one distinctive ideal of political friendship, though they'll have to meet certain other conditions for that too (namely, acceptance of the liberal political values). It doesn't give us any reason to think that this kind of political friendship is non-optional for us, or that we're wronged by the imposition of terms of cooperation that don't accord with the associated ideal. Of course, there's nothing wrong with exploring the ideals internal to a particular kind of relationship. But it's hard to see why people who don't already stand in that relationship should ascribe any fundamental political

importance to the PJR in that case.<sup>29</sup> So on this understanding of the democratic citizenship argument, the pressure to expand the justificatory constituency is avoided, but—as with the stability argument—the moral force of the PJR is radically diminished.

The second of the two options is taken by Paul Weithman's Rawls. He also appeals to the fact that an idealized liberal society is one in which the ideal of democratic citizenship, as specified by the PJR, is upheld. But he avoids the problem I've been describing by presenting this as a reason to expect stability rather than as an independent virtue. The idea is that the ideal is one that citizens of the well-ordered liberal society will tend to see as important and distinct from the value of justice, though clearly closely related (it draws on the same ideals of equality and freedom). (Perhaps they will appreciate the distinctive relationship of political friendship it makes possible, as described in the preceding paragraphs.) So even those who are skeptical about justice as fairness in particular, whose comprehensive views give them grounds for doubt about it, will nevertheless have reason to affirm liberal democratic principles, at least. This is how Weithman (2015b) makes sense of Rawls's continuing to affirm justice as fairness as the best justified conception of justice even as he allows that other liberal conceptions may be legitimate.

---

<sup>29</sup> Compare Gaus's interest in exploring the implicit commitments of the moral "reactive attitudes". If those attitudes really do imply acceptance of a kind of public justification requirement, as Gaus thinks they do, then satisfaction of that requirement does seem to be a fundamental virtue of moral rules. See Gaus (2011: 205–11).

This strikes me as a plausible interpretation. But, since it treats the ideal of democratic citizenship as a reason to expect stability rather than as specifying a fundamental virtue of political cooperation, it clearly embraces the second horn of our dilemma, rather than escaping it. If it's the right interpretation of Rawls, then we should conclude that there is no argument to be found in Rawls's work for any restricted version of the PJR that also packs the kind of moral punch that many contemporary public reason liberals take it to have. And that, indeed, is what I have tried to show.

#### REFERENCES

- Cohen, G.A. 2008. *Rescuing Justice and Equality*. Cambridge, Mass.: Harvard University Press.
- Enoch, David. 2015. Against Public Reason. In *Oxford Studies in Political Philosophy*, volume 1, ed. David Sobel, Peter Vallentyne, and Steven Wall, 112–142. Oxford: Oxford University Press.
- Gaus, Gerald F. 2011. *The Order of Public Reason*. Cambridge: Cambridge University Press.
- Gaus, Gerald F. 2014. The good, the bad, and the ugly: three agent-type challenges to The Order of Public Reason. *Philosophical Studies* 170: 563–577.
- Larmore, Charles. 1999. The Moral Basis of Political Liberalism. *Journal of Philosophy* 96: 599–625.
- Quong, Jonathan. 2011. *Liberalism without Perfection*. Oxford: Oxford University Press.
- Quong, Jonathan. 2013. Public Reason. In *The Stanford Encyclopedia of Philosophy*, summer 2013 edition, ed. Edward N. Zalta. <http://plato.stanford.edu/archives/sum2013/entries/publicreason>.
- Quong, Jonathan. 2014. On the idea of public reason. In *A Companion to Rawls*, ed. Jon Mandle and David Reidy, 265–80. Oxford: Wiley-Blackwell.
- Rawls, John. 1996. *Political Liberalism*. New York: Columbia University Press.
- Rawls, John. 1999. *Collected Papers*, ed. Samuel Freeman. Cambridge, Mass.: Harvard University Press.

- Rawls, John. 2001. *Justice as Fairness: A Restatement*, ed. Erin Kelly. Cambridge, Mass.: Harvard University Press.
- Ripstein, Arthur. 2009. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge, Mass.: Harvard University Press.
- Weithman, Paul. 2010. *Why Political Liberalism?* Oxford: Oxford University Press.
- Weithman, Paul. 2015a. Inclusivism, Stability, Assurance. In *Rawls and Religion*, ed. Tom Bailey and Valentina Gentile, 75–96. New York: Columbia University Press.
- Weithman, Paul. 2015b. Legitimacy and the Project of Rawls's Political Liberalism. In *Rawls's Political Liberalism*, ed. Thom Brooks and Martha C. Nussbaum, 73–112. New York: Columbia University Press.