

Legitimacy and Consensus in Rawls' Political Liberalism

Enzo Rossi

University of Amsterdam

e.rossi@uva.nl

0. Introductory

This paper analyses the theory of legitimacy at the core of Rawls' *Political Liberalism* (PL). There is little doubt that Rawls' later work¹ constitutes the main driving force behind the rise of public justification as the dominant account of political authority in contemporary liberal theory, and so behind the revival of the contract-based liberal tradition.² Rawls' aim in PL is to correct an inconsistency in his celebrated theory of 'justice as fairness'. Rawls' solution to this problem is to change the status of justice as fairness from that of a *comprehensive* moral conception of justice to that of a *political* moral conception. He thus explains why such a change of status is necessary, spells out the normative requirements of a such a *political* conception of justice, and argues that justice as fairness is able to cope with these requirements (perhaps regrettably, he does not always keep these three issues separate). Now the interesting thing here is that these requirements of legitimacy and consistency, according to Rawls, apply to *any* liberal conception of justice (justice as fairness becomes thus just one out of many possible legitimate conceptions of justice), i.e. to the set of normative principles characterizing the political framework. In other words, in PL he intends to provide an account of the foundations of all normative liberal political theory. He does this through a theory of liberal legitimacy whose requirements, in his view, any

¹ I shall focus mainly on *Political Liberalism*, New York, Columbia University Press 1994² (hereinafter PL), but also on the earlier essays that led to the development of this influential volume (*Kantian Constructivism in Moral Theory*, *The Idea of an Overlapping Consensus*, and *The Domain of the Political and Overlapping Consensus*), as well as on Rawls' later restatements and clarifications of his theory of legitimacy through public reason the essay *The Idea of Public Reason Revisited* and the brief monograph *Justice as Fairness: A Restatement*, Cambridge, MA, Belknap Press, 2001. All essays are in J. Rawls, *Collected Papers*, ed. Samuel Freeman, Cambridge, MA, Harvard University Press, 1999.

² I explain why I do not consider *A Theory of Justice* (Cambridge, MA, Harvard University Press, 1971) the starting point of this revival at the beginning of section 1 below.

liberal conception of justice (such as justice as fairness) should satisfy. Here, in presenting the arguments of Pl, I shall isolate the discussion of the normative requirements of liberalism in general, i.e. of liberal legitimacy, from the issue of how justice as fairness copes with them (which takes up a great part of the book), and analyze only the former.

The aim of this paper is to show that Rawls' theory of legitimacy does not satisfy its own desiderata. Rawls argues that a political system is well grounded when it is stable. As I shall make clear, this notion of stability embodies both pragmatic and moral elements, each of which constitutes a key desideratum of Rawlsian liberal legitimacy. My main claim here is Rawls' strategy to overcome the tension between those desiderata through his theory of public justification is ultimately unsuccessful, because the account of consensus it envisages is unstably placed between the extremes of moralized redundancy and pragmatic free-for-all.

The paper is structured as follows. In section 1 I present the main problem that motivates Rawls' quest for a political conception of justice, which I shall call 'the fundamental question'. Then I provide a brief account of the answer to that question, showing how the core of political liberalism is constituted by a theory of legitimacy³, and outline the main tenets of this theory. In section 2 I move on to offer my critique of that theory, which in turn leads to wider considerations on consensus-based theories of legitimacy. In section 3 I summarize my argument and briefly canvass some issues for further investigation.

1. The Structure of Political Liberalism

The above observation that *Political Liberalism* marked the beginning of the revival of the contract-based liberal tradition might sound surprising. After all, Rawls' earlier and most influential work, *A Theory of Justice*, was already characterized by a contractualist approach. Yet that is hardly relevant here, as that work, unlike *Political Liberalism*, was not about legitimacy, but rather about justice. Rawls did not write *A Theory of Justice* in order to provide a *liberal* theory of justice; justice as fairness is a

³ Burton Dreben was among the first interpreters of Rawls to suggest that the theory of legitimacy was the main focus of Pl. See *On Rawls and Political Liberalism*, in S. Freeman (ed.), *The Cambridge Companion to Rawls*, Cambridge, Cambridge University Press, 2003.

theory of justice *simpliciter*. It is only later that Rawls realized that, for reasons we shall explore below, he had to show that justice as fairness could be a legitimate basis for the characterization of a liberal political framework. *Political Liberalism*, then, aims to show precisely that justice as fairness can be understood as being compatible with the values that citizens of a liberal democratic society can reasonably be expected to hold.

To see what motivated that move, consider how, throughout Rawls' work, there remains a very close link between justice and legitimacy: he maintains that we have a natural duty to obey just institutions (so this duty would provide legitimacy to the exercise of authority on the part of just states).⁴ The crucial point is that this duty becomes effective only once we recognize just institutions as such—hence the need for a consensus on the political framework.⁵ As will become clearer shortly, that is why, in Rawls' own words, the main question political liberalism is meant to answer is this: «How is it possible that there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable though incompatible religious, philosophical, and moral doctrines?»⁶ Rawls calls this problem the question of stability. This question sets the agenda of political liberalism, and Rawls' answer to it, then, constitutes the whole normative structure of political liberalism. Let us first get clear about its various components. Rawls' use of the term 'stability' is somewhat misleading, for the question addresses two distinct (and yet connected) issues: the need for an appropriate consensus and the need for peaceful coexistence (or 'stability' in the ordinary sense). We can think of each of these issues as addressing diverging desiderata: a moralistic or voluntaristic one, directed at establishing a free consensus, and a pragmatic or realistic one, directed at including a genuinely wide variety of comprehensive doctrines.

At any rate, before diving into Rawls' terminology, it will be best to start with a working paraphrases of the question, which could be roughly formulated as follows: given that the citizens of liberal democratic societies hold reasonable different and incompatible conceptions of the good life that regulate all aspects of morality, how

⁴ Rawls refers to this duty as 'the duty of justice'. See *A Theory of Justice*, pp. 114ff. and pp. 334-335. Also cf. Pl, pp. 142ff.

⁵ In fact, Rawls observes that «Citizens would not be bound to even a just constitution unless they have accepted and intend to continue to accept its benefits. Moreover this acceptance must be in some appropriate sense voluntary.» (*A Theory of Justice*, p. 336). Also cf. Pl, pp. xviiiif.

⁶ Pl, p. xx.

can they agree on a moral matter such as the principles of justice regulating fundamental aspects⁷ of their life in society? This formulation shows us two important components of the conceptual framework of political liberalism, namely that of reasonable comprehensive doctrine, and that of a political conception of justice regulating the basic structure of society.

A comprehensive doctrine is a conception of the good that covers all or many aspects of what is valuable in human lives, spanning across political, moral, and other values. It can be a religious doctrine, a philosophical doctrine, and the like. As for the reasonableness of a comprehensive doctrine, suffice it for now to say that it is what makes a doctrine worthy of being accommodated in a liberal society.

A political conception of justice is the set of moral rules⁸ that regulate the basic structure of society, that is «society's main political, constitutional, and economic institutions and how they fit together to form a unified scheme of social cooperation from one generation to the next».⁹ That is 'the domain of the political': 'constitutional essentials' and 'basic questions of justice', i.e. what I have been calling 'the political framework'. So a political conception is a moral conception, but not a comprehensive one: it is smaller in scope than a comprehensive conception, for while the former only applies to the domain of the political, the latter regulates all that is valuable in human life.

Now the problem is, how can a plurality of (reasonable) comprehensive doctrines be accommodated by a single political conception of justice? If a doctrine is comprehensive and thus covers all aspects of morality (to consider the hardest case), there cannot be room for a political conception of justice (which is by definition a moral conception) by the side of it. The conception of justice cannot be added, as it were, to the comprehensive doctrine – precisely because the doctrine is comprehensive, all-embracing: it allows for no other authorities in matters of

⁷ Constitutional essentials and matters of basic justice, according to Rawls.

⁸ Here (following Rawls) I am using 'moral' in a rather loose sense, according to which 'moral' rules are articulated on the basis of certain 'values'. In the case of a political conception of justice, they are 'political values' (PI, p. 11*n*). It is not very clear whether Rawls wants to draw a line between prudential, moral and other types of values, so I will assume a rather capacious notion of 'value', one comprising any ideal or principle informing one's conception of the good life.

⁹ PI, p. 11. The definition of the *concept* (as opposed to a *conception*) of justice from *A Theory of Justice* could appropriately be employed here: «A characteristic set of principles for assigning basic rights and duties and for determining what they take to be the proper distribution of the benefits and burdens of social cooperation». (*A Theory of Justice*, p. 5). The only difference, in PI, is that the scope of justice is narrower and limited to 'the domain of the political'. It is correct to employ this definition because here we are talking about a *conception* of justice.

morality, and for no other subjects besides those it covers. So, if they are to co-exist, the two doctrines must coincide, or overlap. That is to say, the political conception of justice must be part of the comprehensive doctrine – the problem being that, while there is a plurality of comprehensive doctrines, for society to be well-ordered there must be just one political conception of justice:

...a continuing shared understanding on one comprehensive religious, philosophical or moral doctrine can be maintained only by the oppressive use of state power. If we think of political society as a community united in affirming one and the same comprehensive doctrine, then the oppressive use of state power is necessary for political community. [...] Call this 'the fact of oppression'.¹⁰

The sort of oppression Rawls has in mind here has to do with what one may call 'anti-voluntarism': a legitimate liberal state must strive to at least approximate voluntary allegiance on the part of its citizens:

We may over the course of life come freely to accept, as the outcome of reflective thought and reasoned judgment, the ideals, principles, and standards that specify our basic rights and liberties, and effectively guide and moderate the political power to which we are subject. This is the outer limit of our freedom.¹¹

How can we reach the outer limits of our (political) freedom? The problem of achieving convergence without oppression is dealt with through the ideas of (i) 'political constructivism' and (ii) 'overlapping consensus': a political conception of justice (i) must be *constructed as if* it had been derived from *any* reasonable conception of the good, so that (ii) it will be capable of becoming the object of an *overlapping consensus* of reasonable comprehensive doctrines. In other words, the political conception of justice must be *modular*: it must be such that it could be grounded in

¹⁰ Pl, p. 37.

¹¹ Pl. p. 222.

any of the numerous reasonable comprehensive doctrines. And for this modularity to be possible, the political conception must be presented as *freestanding*.¹²

While we want a political conception to have a justification by reference to one or more comprehensive doctrines, it is neither presented as, nor as derived from, such a doctrine applied to the basic structure of society, as if this structure were simply another subject to which that doctrine applied. [...] We must distinguish between how a political conception is presented and its being part of, or as derivable within, a comprehensive doctrine. I assume all citizens to affirm a comprehensive doctrine to which the political conception they accept is somehow related. But a distinguishing feature of a political conception is that it is presented as freestanding and expounded apart from, or without reference to, any such wider background.¹³

That is to say that the formulation of the political conception must not depend on any comprehensive doctrine, for that would make it unacceptable to proponents of other doctrines. Only in this way can an overlapping consensus obtain, and it is also in the light of the idea of an overlapping consensus that the political conception is neutral with regard to the good, in (at least) two senses: (i) the basic structure of society as defined by the political conception should not *aim* at favoring any particular comprehensive doctrine; (ii) as the focus of an overlapping consensus, the political conception constitutes common ground—i.e. neutral ground—between different and diverging comprehensive doctrines.¹⁴ The idea of neutrality, together with that of the priority of the right over the good, should not be understood as completely forbidding the use of non-instrumental ideas of the good. Rawls does indeed acknowledge that «the right and the good are complementary: no conception of justice can draw entirely upon one or the other, but must combine both in a

¹² It may be interesting to note how this idea had been anticipated by Judith Shklar: «Liberalism does not in principle have to depend on specific religious or philosophical systems of thought. It does not have to choose among them as long as they do not reject toleration.» (*The Liberalism of Fear*, in N. Rosenblum (ed.), *Liberalism and the Moral Life*, Cambridge, MA, Harvard University Press, p. 24).

¹³ Pl, p. 12.

¹⁴ See Pl, pp. 192-194.

definite way.»¹⁵ The point of liberal neutrality is rather that the ideas of the good employed by the political conception of justice must be presented in such a way that does not commit the conception of justice to affirming the truth of one particular comprehensive doctrine. In other words, the ideas of the good must be *political* ideas of the good, i.e. limited to subjects pertaining to the domain of the political.¹⁶ Also one should of course not take ‘neutrality about the good’ to mean ‘neutrality about morality’: the political conception of justice is a moral conception. The relevant opposition, here, is political *vs* comprehensive, not political *vs* moral.

The challenge of political liberalism, then, is to explain exactly how such an overlapping consensus can take place. In other words, how one could possibly maintain a liberal understanding of politics by the side of a non-liberal (religious, philosophical, etc.) understanding of the good life. That problem will not be our main focus here—at least not directly. Let me just mention very briefly, for the sake of giving an idea of the broader picture of Rawls’ theory, what is the strategy that enables a political conception of justice to become the focus of an overlapping consensus of reasonable comprehensive doctrines. Roughly put, for the overlapping consensus to be possible, the political conception (and thus the debates about its subject: ‘constitutional essentials’ and ‘questions of basic justice’) must not employ values and ideas that are only intelligible or acceptable to proponents of a comprehensive doctrine about which reasonable people can be expected to disagree. It must instead be publicly justified, i.e. expressed in terms of values and ideas implicit in the public culture of society; in Rawls’ terminology, it must be expressed in compliance with *public reason*. Are these reasons, in virtue of their being implicit in the public culture, available to citizens as they actually are, or as they should be? Put another way, is this a primarily realistic or idealistic conception of public reason? The latter reading seems to be the most plausible one. These ideas may very well be

¹⁵ *Ibid.*, p. 173.

¹⁶ Thus Rawls says that «The priority of right means that the ideas of the good used must be political ideas [...], so that we need not rely on comprehensive conceptions of the good but only on ideas tailored to fit within the political conception.» (PI, p. 209). It is hardly possible to discuss the plausibility of this notion of political ideas of the good in abstract, without reference to a particular conception of justice, for each conception of justice will employ different ideas of the good. Rawls discusses the case of the five ideas of the good that can be identified in justice as fairness in (PI, Lecture V). Various different uses of the term ‘neutrality’ are discussed in PI, pp. 190-195. Another useful discussion of the various senses of ‘neutrality’ (which, just like Rawls’, is largely derived from the work of Ronald Dworkin) is in Charles Larmore’s *The Morals of Modernity*, Cambridge, Cambridge University Press, 1998, pp. 125-127.

implicit in the public culture, but that does not mean that they are available to all citizens, taken as they are. Indeed, as we shall see below, they are only available to 'reasonable' citizens.

Whether or not such an overlapping consensus is possible, and whether or not the conception of justice at its focus can be true to the requisites of neutrality are the issues that have generated the great majority of the criticisms of the project of political liberalism.¹⁷ That is in fact a vital problem for the assessment of political liberalism, but it is not the one I am interested in here. So I shall grant Rawls that there is a sense in which the overlapping consensus is possible; what I am interested in, then, is exactly what this sense is. Indeed, my concern is whether this possibility can suffice to satisfy the desiderata of a liberal theory of legitimacy through public justification.

What I have tried to show so far is in fact just that the normative structure of political liberalism is designed in order to accommodate reasonable pluralism. We can say that the structure of political liberalism is a response to the challenge of diversity, i.e. the problem, faced by liberal legitimacy theorists, of providing an account of how citizens may converge on a political framework in a society characterized by the diversity brought about by the legitimacy theorists' preferred form of political organization. The conception of justice, for instance, needs to be political and not comprehensive precisely because the doctrines held by the citizens are comprehensive and plural. And the 'politicality' of a conception of justice is what makes it justifiable. In other words, the question of stability comes into play and sets the agenda of political liberalism only once we take reasonable pluralism seriously, i.e. when we realize that it is appropriate and necessary for a conception of justice to accommodate a plurality of incompatible reasonable comprehensive doctrines.

A question arises quite naturally here: we know from our account of the challenge of diversity that a liberal political framework should seek to accommodate ethical diversity in order to secure the consensus of the citizenry, however—in the case of neutralistic political liberalism—why accommodate precisely the measure of ethical diversity defined through the idea of reasonable comprehensive doctrines? As

¹⁷ Clearly another main issue of the debate on Rawls' political liberalism has been the ability of the theory of justice as fairness can abide by the standards of political legitimacy set out in Pl. This issue is in fact the main concern of Pl as intended by Rawls. Here however, as I stated in the introductory section, I am concerned only with the framework of requirements of political legitimacy, and not with its application to any particular conception of justice (such as justice as fairness).

we shall see, answering this question will reveal a crucial aspect of the theoretical structure Rawls deploys in order to establish the possibility of an overlapping consensus.

However, before turning directly to that problem, it will be useful to look at political liberalism's fundamental question more closely. At the beginning of this section I anticipated that the fundamental question, the question of stability, can be understood as resulting from the combination of two problems, namely the problem of stability and that of legitimacy. Alternatively, and perhaps more precisely, we may say that *Political Liberalism's* account of legitimacy incorporates an account of stability. In the next section, then, I shall explain why this is so, and how. That is to say, I shall provide an account of both issues, and show how they are connected to each other in the fundamental question.

2. Stability for the Right Reasons

It is now time to understand what Rawls means when he says that the fundamental question of political liberalism is that of 'stability'. There is an intuitive sense in which the fundamental question concerns stability (in the ordinary sense of the term), in that it asks about the 'existence over time' of a 'stable society' of citizens 'profoundly divided by incompatible moral doctrines'. One might in fact think that stability is just about making society's institutions persist in spite of conflicts between ethically diverse citizens. This familiar sense of 'stability', however, is not Rawls'—indeed, he does from time to time specify that he is interested in a special kind of stability, which he refers to as «stability for the right reasons»¹⁸. But then what exactly does 'stability' mean in the context of political liberalism? Strictly speaking stability has to do with the motivation, on the part of the citizens, to comply with the conception of justice regulating the basic structure of society. More precisely, the general problem of stability concerns how citizens can develop and maintain a sense of justice such that they will be motivated to comply with the conception of justice regulating the basic structure.¹⁹ What is important to point out here is that, as noted above, the stability required by the standards of justification of political liberalism is not merely—or perhaps even primarily—a pragmatic matter. It does not have to do just

¹⁸ See PI, pp. xlii, 388n, 390, 392.

¹⁹ *Ibid.*, pp. 141ff.

with the pragmatic problem of securing social order and avoiding the outburst of open, perhaps violent conflicts between individuals or groups divided by incompatible comprehensive doctrines.²⁰ (A social order supported only by a necessarily contingent equilibrium of power between conflicting comprehensive doctrines is termed by Rawls a ‘modus vivendi’.) Hence the sense of justice that should cause the citizens to honor the political conception of justice is not to be understood as stemming just from prudential considerations, but rather from moral ones. But why is prudence-based, modus vivendi-stability not enough?²¹ Answering this question will lead us towards the second issue implicit in the fundamental question of political liberalism, legitimacy, and it will show the connection between both issues.

Modus vivendi-stability is not acceptable because it would violate a requirement that any liberal political conception of justice must comply with. This requirement is what Rawls calls ‘the liberal principle of legitimacy’. The liberal principle of legitimacy, of course, is a moral principle. The thought behind the liberal principle of legitimacy is quite familiar. In a liberal democratic regime the citizens are viewed as free and equal, and political power is exercised by them jointly, precisely in virtue of their freedom and equality. Hence any use of political power that is not presented in a way that citizens could be reasonably expected to endorse ‘for the right reasons’ does not respect their freedom and equality. To repeat, the citizens’ freedom and equality are the only legitimate sources of political authority, and the use of political power must, as it were, be accountable to them.²² Thus the liberal principle of legitimacy is a moral principle in the sense that it stems from a moral conception of persons as free and equal and as possessing the two moral powers of

²⁰ It is important to stress this point because the formulation of the question of stability for political liberalism has led many commentators to see a Hobbesian Rawls, seeking just the avoidance of conflict.

²¹ On the moral differences between overlapping consensus and modus vivendi see my *Modus Vivendi, Consensus, and (Realist) Liberal Legitimacy*, in «Public Reason», 2 (2010), no. 2, pp. 21-39, and *Consensus, Compromise, Justice and Legitimacy*, in «Critical Review of International Social and Political Philosophy», 16 (2013), no. 4, pp. 557-572.

²² As clarified by the quotation below, this does not mean that every political decision must be such that citizens could be expected to directly endorse it (that would leave no room at all for the democratic confrontation of different political outlooks). This endorsement is however granted indirectly through the endorsement of the *constitutional* system that makes every political decision possible and legitimate.

forming and exercising a sense of justice and a conception of the good.²³ (As it should be clear, in Rawls' theory this conception performs the function of what in the previous paper I vaguely referred to as the value of personal autonomy: it provides the premises from which the appeal of the legitimacy view flows.) Rawls formulates the principle as follows:

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. This is the liberal principle of legitimacy.²⁴

Rawls also refers to this principle as 'criterion of reciprocity':

«By what ideals and principles, then, are citizens as sharing equally in ultimate political power to exercise that power so that each of them can reasonably justify their political decisions to each other?»

«The answer is given by the criterion of reciprocity: our exercise of political power is proper only when we sincerely believe that the reasons we offer for our political action may reasonably be accepted by other citizens as justification of those actions.»²⁵

In a footnote, then, Rawls explains that this criterion of reciprocity «is a *duty* arising from the idea of reasonableness of persons...».²⁶ The fact that the idea of reasonableness can generate duties indicates its moral connotation.²⁷ The idea here is that if a liberal democratic state wants to take its citizens' freedom and equality

²³ This conception of the person must in turn be presented as a purely political one, i.e. not dependent on any controversial metaphysical doctrine and thus publicly intelligible for reasonable and rational citizens. In Pl, pp. 29-35 Rawls explains in some detail how the conception of the person employed in justice as fairness can be presented in such a way. The two moral powers and the way they help defining the conception of the citizens as free and equal are discussed specifically in Pl, pp. 81ff.

²⁴ Pl, p. 137.

²⁵ Pl, p. xlvi.

²⁶ *Ibid.*, xlviii, my italics

²⁷ I shall come back to this rather important point in the next section. On the morality of the reasonable see S. Mulhall and D. Swift, *Liberals and Communitarians*, London, Wiley Blackwell, 1996², pp. 231ff.

seriously, its political life must be regulated by a constitution upon which the citizens can freely agree, as equals. That rules out the *modus vivendi* constitutional agreement because it does not stem from a freely endorsed moral conception of justice, but rather from the necessity of preventing open conflict (in a sense, choosing a *modus vivendi* arrangement is like ‘choosing’ to give up one’s wallet at gunpoint). Given then that a *modus vivendi* would be illegitimate from a liberal point of view, we need to find a moral conception of justice. (I take it that, as Locke suggested with regard to religious doctrines, a moral conception is something one can only endorse freely or not endorse at all—one can of course be forced to act on an unendorsed moral conception, but that is precisely what the liberal principle of legitimacy tries to avoid as much as possible.)²⁸

Now the problem, as we know, is indeed that the moral conception of political justice is part of one’s comprehensive doctrine, but there is no agreement, in liberal societies, on comprehensive doctrines. Combining the liberal ideal of legitimacy with the fact of reasonable pluralism, then, we arrive at the formulation of the question of stability we saw at the outset of our discussion: «How is it possible that there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable though incompatible religious, philosophical, and moral doctrines?»²⁹ At this point we already know Rawls’ answer to the question, but it is important to reflect on it in order to shed light on two related matters. Firstly, as it should become clear shortly, this question is Rawls’ way of introducing his strategy for responding to what I have labeled the challenge of diversity. Secondly, the combination of the liberal principle of legitimacy and of the question of stability in the light of the fact of reasonable pluralism makes the question of stability a moral question, at least in part. In fact, now that we are acquainted with Rawls’ liberal principle of legitimacy and with its connection to the problem of stability, we can see how the fundamental question of political liberalism does not just ask whether people can peacefully co-exist notwithstanding their divisions on matters of the good life (a *modus-vivendi* arrangement would be enough to answer this question affirmatively). He mentions

²⁸ ‘As much as possible’ in the sense that the principle will nonetheless sanction the imposition of a moral conception upon unreasonable citizens. As Rawls puts it, unreasonable doctrines should be ‘contained’ (PI, p. xix). Jonathan Quong elaborates on this point in *The Rights of Unreasonable Citizens*, in «Journal of Political Philosophy», 12(2004), no. 3 pp. 314-335.

²⁹ *Ibid.*, p. xx.

instead a «just society of free and equal citizens», thus referring to the liberal political moral conception of the person the liberal principle of legitimacy is meant to protect:

The kind of stability required of justice as fairness [and thus of any other legitimate liberal conception of justice] is based, then, on its being a liberal political view, one that aims at being acceptable to citizens as reasonable and rational as well as free and equal, and so as addressed to their public reason. [...] If justice as fairness were not expressly designed to gain the reasoned support of citizens who affirm reasonable although conflicting comprehensive doctrines—the existence of such comprehensive doctrines being a feature of the kind of public culture that liberal conception itself encourages—it would not be liberal.³⁰

And to be liberal (i.e., for Rawls, to comply with the liberal principle of legitimacy), we have seen, is a moral matter. To sum up: the question of stability is (also) a moral question because its formulation depends on a moral conception of the citizens as free and equal that is at the core of any liberal theory of legitimacy (it is in this sense that Rawls sometimes uses the phrase «stability for the right reasons»). In other words, political liberalism must respond to the question of stability and thus accommodate reasonable pluralism through a shared moral political conception of justice because imposing on free and equal citizens a conception of justice they cannot endorse would fail the test of liberal legitimacy. However, it is important to point out how the question is not purely a moral one: there is a pragmatic aspect to it, for the very form of the question shows that Rawls wants to argue that it is possible for «a stable and just society of free and equal citizens profoundly divided by reasonable though incompatible religious, philosophical, and moral doctrines» to exist and persist.³¹ This aspect is driven by what one may call the pragmatic desideratum of a theory of liberal legitimacy. Below I shall come back to this issue, which will play a crucial role in my critique of the Rawlsian project; in order to see why this is the case, however, we need to get clear about another aspect of Rawls' theory, namely his notion of *reasonableness*. Indeed, reasonableness plays a very

³⁰ Pl, p. 143.

³¹ I take it that what Rawls has in mind is not mere logical possibility, which does not seem enough in the context of normative political theory.

important role in the structure of political liberalism's theory of legitimacy, because, as we shall see shortly, it ensures that the agreement obtained through an overlapping consensus will meet the requirements posed by the theory's moralistic desiderata: it ensures that the agreement will be an agreement between free and equal citizens, i.e. an agreement that preserves personal autonomy, the foundational commitment we have identified as the central moral commitment of public justification. In other words, by shaping the rules for the conduct of public deliberation (for only reasonable citizens may take part in the procedure), reasonableness specifies what it takes for consent to qualify as genuinely free consent.

Analyzing the content of the notion of reasonableness brings us back to the question we left open at the end of the previous section: why does a liberal political conception of justice have to be designed in such a way that it can accommodate ethical diversity to the extent that it is *reasonable*? It could appear that the question has been answered in the previous section, in so far as we have seen the connection between the question of stability, reasonable pluralism, and the liberal principle of legitimacy. Actually, to some extent it has been answered. Reversing the argument we have just seen, in fact, we could say that reasonable pluralism should be accommodated in virtue of its being the product of the exercise of the free reason of the citizens: if, as good liberals should do, we want to take the citizen's freedom and equality seriously, we have to respect the outcome of the free exercise of their moral powers,³² namely reasonable pluralism. This might be an answer to our question. However, on closer inspection, it does not seem a very satisfactory one. In fact, it does not explain precisely in what sense *reasonable* pluralism is to be accommodated as an instantiation, as it were, of the freedom and equality of the citizens, whereas 'unreasonable' comprehensive doctrines should not be tolerated. In other words, despite the fact that any conception of the good actually existing in a liberal society is produced by the free exercise of human reason, only the reasonable ones ought to be accommodated in the liberal polity. Why is this? Clearly in order to answer the question properly we must analyze the notion of reasonableness. We start by looking at how Rawls characterizes reasonable comprehensive doctrines:

³² On the two moral powers of the citizens see Pl, pp. 81ff.

This fact of reasonable pluralism must be distinguished from the fact of pluralism as such [...]. [Reasonable comprehensive doctrines] are not simply the upshot of self- and class interest, or of people's understandable tendency to view the political world from a limited standpoint.³³

This points towards the idea that, as noted above, a reasonable doctrine should not be accommodated simply because it is expressed by a free and equal citizen under free institutions, but rather because there is something in it, *qua reasonable*, that makes it more valuable than just any conception of the good that might be conceived under free institutions. The value of the doctrine, what makes it worthy of accommodation in the liberal polity, is its reasonableness. In fact, Rawls also refers to reasonableness as a political virtue that grounds liberal toleration:

...The idea of the reasonable especially, whether applied to persons, institutions or doctrines, easily becomes vague and obscure. I try to mitigate this by fixing on two basic aspects of *the reasonable as a virtue of persons engaged in social cooperation among equals*. I then develop from those two aspects the content of the reasonable. I examine next how doing this provides a basis for toleration in a society marked by reasonable pluralism.³⁴

But what exactly does this reasonableness consist in? What is its connection with liberal toleration? In order to answer these questions, let us have a quick look at those two key aspects of reasonableness. They are (i) «the willingness to propose fair terms of cooperation and to abide by them provided others do», and (ii) «the willingness to recognize the burdens of judgment and to accept their consequences for the use of public reason in directing the legitimate exercise of political power in a constitutional regime.»³⁵

So, *first*, a reasonable person will not be guided and motivated just by the pursuit of his or her own self-interest, but will also possess «the particular form of

³³ *Ibid.*, p. 37.

³⁴ *Ibid.*, 48, my italics.

³⁵ *Ibid.*, p. 54.

moral sensibility that underlies the desire to engage in fair cooperation as such, and to do so on terms that others as equals might be reasonably expected to endorse».³⁶ This feature of reasonableness is connected to the citizen's moral power of having a sense of justice,³⁷ and builds into reasonableness a moral commitment that, as we will see below, ties it up with core values of political liberalism: «Reasonable persons [...] desire for its own sake a social world in which they, as free and equal, can cooperate with others on terms all can accept.»³⁸

In order to understand the *second* point we need to get clear about the notion of 'the burdens of judgment.' The burdens of judgment are introduced in order to account for how reasonable people can fail to reach an agreement about a comprehensive moral doctrine. These burdens are «the many hazards involved in the correct (and conscientious) exercise of our powers of reason and judgment in the ordinary course of political life.»³⁹ The important point here is that reasonable citizens, in acknowledging the burdens of judgment, realize that free cooperation between equals—which is required by (i)—is only possible if political discourse (at least when constitutional essentials and matters of basic justice are at stake) is conducted in terms that are intelligible and acceptable to every other reasonable citizen: «reasonable persons see that the burdens of judgment set limits on what can be reasonably justified to others, and so they endorse some form of liberty of conscience and freedom of thought.»⁴⁰

These points show us not only that the notion of reasonableness has a strong moral connotation, but also that it is connected with distinctively liberal conceptions of the person and of society:

Observe that here being reasonable is not an epistemological idea (though it has epistemological elements).⁴¹ Rather, it is part of a political

³⁶ *Ibid.*, p. 51.

³⁷ *Ibid.*, p. 52.

³⁸ With regard to this, Mulhall and Swift noted that «on Rawls' understanding of the term, no one can be reasonable unless they accept the conception of the person and of society that is the irreducible core of political liberalism.» (*Liberals and Communitarians*, p. 236).

³⁹ These sources of disagreement are listed in PI, pp. 56-57.

⁴⁰ *Ibid.*, p. 61.

⁴¹ Another point that makes clear why the concepts of reasonableness and of the burdens of judgment are characterized more morally than epistemological (though they do feature epistemological elements) is given by Rawls' discussion of why the overlapping consensus of reasonable doctrines cannot be grounded on some form of skepticism about morality and the good, for skepticism is too

ideal of democratic citizenship that includes the idea of public reason. The content of this ideal includes what free and equal citizens as reasonable can require of each other with respect to their reasonable comprehensive views.⁴²

Now we see why political liberalism must accommodate *reasonable* comprehensive doctrines: they are doctrines that share a liberal understanding of politics, an understanding of politics involving the idea of fair cooperation and, most importantly, the idea of free and equal citizenship. In other words, every reasonable person, on Rawls' notion of reasonableness, is a political liberal. Hence clearly political liberalism would be self-defeating were it not able to accommodate all reasonable persons (and their doctrines). But political liberalism does not have this problem, for the notion of reasonableness is constructed in such a way that reasonable people will in fact be motivated to seek an overlapping consensus between their comprehensive doctrines. It is the very definition of reasonableness, in fact, that requires them to be committed to seek fair terms of cooperation between free and equals, and that makes them realize (through the acknowledgment of the burdens of judgment) that such terms of cooperation cannot be based on a shared comprehensive moral doctrine. In other words, it is the very definition of reasonableness that commits them to political liberalism. So the answer to what we termed the fundamental question, «how is it possible that there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable though incompatible religious, philosophical, and moral doctrines?», becomes simple: it is possible, provided that we define these reasonable doctrines, qua reasonable, as committed to «a stable and just society of free and equal citizens».

It now becomes clear why the notion of reasonableness is deployed: the idea is that reasonable citizens, in virtue of their commitment to treating other as free and equals, will not agree on a conception of justice regulating the political framework which violates the personal autonomy of their fellow citizens. However, clearly

controversial a doctrine for serving as a basis for such consensus (*Ibid.*, pp. 150-154; on how this relates to the burdens of judgment see pp. 62-63).

⁴² *Ibid.*, pp. 61-62.

someone's autonomy will be violated, namely that of unreasonable citizens.⁴³ The challenge here for Rawls, then, is to justify this exclusion in a non-arbitrary way, i.e. he needs to account for the reasons why a *determinate* subset of the citizenry are included in the consensus, and determinate others are not. It is not clear to me that he succeeds in this task, or indeed that he can succeed.⁴⁴ On my reading of his theory, Rawls could tackle that challenge in two ways: (i) he could argue that the grounds for excluding the unreasonable citizens are given by the fact that otherwise there would be no agreement, or that (ii) his particular account of reasonableness is the only one that ensures genuinely free consent. Yet both lines of argument are problematic. In the case of (i), note how Rawls attempts to justify a failure in fully meeting the moralistic desideratum by appeal to the need of satisfying the pragmatic desideratum, too. Reasonableness, he may say, is deployed in order to identify the largest pool possible of citizens who can consent (at least hypothetically) to a political framework without forfeiting their personal autonomy. It seems to me that that would be a good answer to a pragmatic problem, but it does not cut any ice as far as the moral problem is concerned. What is perhaps more, it brings us back to the worry about the arbitrariness of the pairing of public justification's two desiderata: even if we argue that the pairing is necessary for the sake of securing an agreement, we are still short of an argument for why securing an agreement should be our primary concern. As for (ii), the problem is that, if Rawls went down that route, he would have to commit to the view that consent can only be free when expressed in favor of a liberal political framework, given that we have seen that to be reasonable,

⁴³ It is important to note that it has been convincingly argued that unreasonable citizens make up a far greater proportion of the citizenry even of liberal democratic societies than Rawls seems prepared to acknowledge. For example, while Rawls maintains that most mainstream, non-fundamentalist Christian outlooks are reasonable comprehensive doctrines, Leif Wenar has shown that this is not the case with the Roman Catholic religion: see *Political Liberalism: An Internal Critique*, in «Ethics» 102 (1995), no. 1, pp. 32-62. This example is important because it would probably be excessively demanding to require that every single citizen should be included in the consensus: surely there are uncontroversially unreasonable citizens, in the ordinary sense of the term. But, as Wenar's work shows, Rawls' account of reasonableness rules out conceptions of the good that are widely held and commonly regarded as mainstream.

⁴⁴ Jonathan Quong (*Liberalism Without Perfection*, Oxford: Oxford University Press, 2011, pp. 137–139) defends political liberalism from this sort of objection by arguing that Rawls' theory of legitimacy is aimed merely at liberal self-understanding – a plausible interpretation, yet one that would please many enemies of liberalism, at least insofar as political liberalism seems to want to exclude more robust advocacy of liberal normative commitments.

for Rawls, equals being committed to upholding a liberal political framework.⁴⁵ But that move makes the search for consensus look rather redundant.

Here Rawls may be tempted to reply that his theory sits in a comfortable middle ground between these two pitfalls. However, in order to be able to say that, he should demonstrate that the boundaries of reasonableness have not been set arbitrarily (in this sense I said, at the end of the previous section, that Rawls should justify the exclusion of unreasonable doctrines in a way that does not depend on his pre-existing commitment to liberal values as such), but there is no trace of such an argument in PI or in his most recent work on these issues. Thus, in a society characterized by a level of persistent ethical diversity such as that present in contemporary Western societies, there are little if any hopes to ground the exercise of political authority through anything like political liberalism's theory of legitimacy.

In the light of those points, a further observation is worth making, in order to reinforce the worries I advanced so far. Imagine (counterfactually) that the society addressed by political liberalism's theory of legitimacy were ethically homogeneous enough for us to be able to say that a sufficient percentage of the citizens is reasonable. In this case, we may say that, on public justification, the political framework is indeed legitimate. But the important question would then be: does the (hypothetical) consent of the citizenry carry out at least part of the task of legitimating liberal political authority, or is that job done by the fact the normative constraints imposed on the consensus through the notion of reasonableness ensure that the agreement will be respectful of the core liberal axiological commitments of freedom, equality, and autonomy? Indeed, a close inspection of the theory reveals that the fact that the citizens happen to be reasonable, and thus hypothetically consent to the political framework in Rawls' intended way, is entirely contingent. What does the work needed to ground political authority, then, are the liberal values embedded in the notion of reasonableness, rather than the fact that citizens consent. So it seems that, while Rawls' theory's structure requires a form of consensus voluntaristic enough to carry some independent normative force, the form of hypothetical consent through public justification it envisages hardly meets that

⁴⁵ It is this 'positive' aspect of Kantian-derived conceptions of freedom and autonomy that made many early liberals highly suspicious of Kant (See Raymond Geuss, *Liberalism and Its Discontents* in Id., *Outside Ethics*, Princeton, NJ, Princeton University Press, 2005, pp. 16-17). In fact, it is only after Rawls' contributions that it has become common to count Kant among the key exponents of the liberal tradition.

requirement. That is because it may very well be the case that citizens happen to hypothetically consent to the political system, but whether they do or not is indifferent, as far as the theory is concerned. A Rawlsian may reply that consensus and reasonableness are two necessary conditions for legitimacy, whose conjunction is necessary and sufficient. But that hardly assuages our worries here, since each of those features have to be freestanding in order to have normative force (for instance, as noted above, why bother with requiring consent if it only counts when it comes down in favor of a predetermined outcome?).

Finally, it may be interesting to note how my proposed reading and criticism of PI may provide an insight into one controversy among interpreters and critics of Rawls' later work (hopefully this should also clarify my own interpretation). In fact, it is possible to classify some readers of PI as favoring a moralistic interpretation, and others as favoring a pragmatic one. Larry Krasnoff constitutes one of the best examples of the former approach, whereas Leif Wenar and Jürgen Habermas are quite representative of the latter, albeit in rather different ways.⁴⁶ Now, if my interpretation is correct, it follows that there is some truth on both sides of the debate, and that it is Rawls' own failure to resolve the tension between the moralistic and the pragmatic aspirations of his theory that could explain why some of his readers, perhaps in an attempt to identify a more coherent theory than what Rawls actually offered, have been led to disregard the one or the other of the two vital and yet fatal aspects of Rawls' theory.⁴⁷

3. Conclusion

We started out by observing how the core business—as it were—of Rawls' *Political Liberalism* is to present a theory of legitimacy for liberal polities, and that this particular theory is indeed the blueprint for contemporary liberal legitimacy theory. We have seen that Rawls' reason for putting forward such a theory is that even a just state acquires full legitimacy only if it is consented-to by its citizenry—a consensus

⁴⁶ See L. Krasnoff, *Consensus, Stability and Normativity in Rawls' Political Liberalism*, in «Journal of Philosophy», 95 (1998), no. 6, pp. 269-292; L. Wenar, *Political Liberalism: An Internal Critique*, J. Habermas, *Reconciliation Through the Public Use of Reason: Remarks on John Rawls' Political Liberalism*, in «Journal of Philosophy», 92 (1995), no. 3, pp. 132-180.

⁴⁷ For a development of this issue in light of the recent debate on realism in political theory, see F. Freyenhagen, *Taking Reasonable Pluralism Seriously*, in «Politics, Philosophy & Economics», 10 (2011), no. 3, pp. 323-342, J. Gledhill, *Rawls and Realism*, in «Social Theory and Practice», 38 (2012), no. 1, pp. 55-82, and R. Jubb, *Playing Kant at the Court of King Arthur*, forthcoming in «Political Studies» (2014).

which, in Rawls' view, is achieved through the idea of public justification. The main problem in securing such a consensus, though, emerges once we realize that the freedom afforded by liberal democratic societies leads to the emergence and persistence of wide ethical diversity. The idea of public reason is then introduced in order to devise a standpoint from which the political framework can be seen as justified in terms of values present in the public culture of such a society—that is done through an account of the rules for public deliberation that would have had to be followed to arrive at the presently justified political framework, and that will have to be followed when debating changes to it. Yet these rules also have to ensure that the consensus obtained is free in a relevant way, i.e. that the personal autonomy of the citizens will be safeguarded throughout the deliberation process. Thus Rawls introduces the notion of reasonableness: only reasonable citizens may take part in the deliberation, and to be reasonable means to be committed to respecting the freedom and equality of other citizens. Here, however, we have noted how this account of reasonableness is problematic, as it arbitrarily builds liberal outcomes into the deliberation procedure—for what is it to be committed to the freedom and equality of one's fellow citizens if not to be a liberal? Now that is problematic because there is no argument within Rawls' theory accounting for why we can place exactly those constraints on public reason. In other words, either it is not clear why only consent to liberal institutions and policies produces legitimacy, or 'free consent' is constructed in such a way as to result genuinely free only when expressed in favor of liberal institutions and policies, thus becoming rather redundant. I have tried to show how that disjunction originates from a problematic tension between two desiderata of political liberalism: a moralistic or voluntaristic one (seeking consensus), and a pragmatic or realistic one (seeking inclusivity). Other theories of public justification may succeed in striking a better balance between those desiderata; yet this paper's argument gives us some reason to consider the problem a structural defect of the very idea of public justification,⁴⁸ rather than simply of Rawls' version of it.⁴⁹

⁴⁸ For a discussion of how those worries generalize to other theories of legitimacy through public justification see J. Bohman and D. Richardson, *Liberalism, Deliberative Democracy, and 'Reasons That All Can Accept'*, in «Journal of Political Philosophy», 17 (2009), no. 3, pp. 253-274, and my *Legitimacy, Democracy, and Public Justification: Rawls' Political Liberalism vs Gaus' Justificatory Liberalism*, in «Res Publica», Online First (2013) doi:10.1007/s11158-013-9223-9.

⁴⁹ Versions of this paper were presented at the Universities of Manchester, Pavia, and St Andrews. I'm grateful for the comments received from all audiences for their comments, as well as from Rowan Cruft, Tim Mulgan, John Skorupski, Leif Wenar and two anonymous referees for *Iride*. This piece

References

Bohman, J. and D. Richardson, Liberalism, Deliberative Democracy, and 'Reasons That All Can Accept', in «Journal of Political Philosophy», 17 (2009), no. 3, pp. 253-274.

Dreben, B., On Rawls and Political Liberalism, in S. Freeman (ed.), *The Cambridge Companion to Rawls*, Cambridge, Cambridge University Press, 2003.

Freyenhagen, F., Taking Reasonable Pluralism Seriously, in «Politics, Philosophy & Economics», 10 (2011), no. 3, pp. 323-342.

Geuss, R., *Liberalism and Its Discontents* in Id., *Outside Ethics*, Princeton, NJ, Princeton University Press, 2005

Gledhill, J. Rawls and Realism, in «Social Theory and Practice», 38 (2012), no. 1, pp. 55-82.

Habermas, J., Reconciliation Through the Public Use of Reason: Remarks on John Rawls' Political Liberalism, in «Journal of Philosophy», 92 (1995), no. 3, pp. 132-180.

Jubb, R., Playing Kant at the Court of King Arthur, forthcoming in «Political Studies» (2014).

Krassnoff, L., Consensus, Stability and Normativity in Rawls' Political Liberalism, in «Journal of Philosophy», 95 (1998), no. 6, pp. 269-292;

Larmore, C., *The Morals of Modernity*, Cambridge, Cambridge University Press, 1998.

Mulhall, S. and D. Swift, *Liberals and Communitarians*, London, Wiley Blackwell, 1996

Quong, J., *Liberalism Without Perfection*, Oxford: Oxford University Press, 2011.

Quong, J., The Rights of Unreasonable Citizens, in «*Journal of Political Philosophy*», 12(2004), no. 3 pp. 314-335.

Rawls, J. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.

Rawls, J., *Political Liberalism*, (2nd ed.) New York: Columbia University Press, 1994.

Rawls, J., *Collected Papers*, Cambridge, MA: Harvard University Press.

Rawls, J., *Justice as Fairness: A Restatement*. Cambridge, MA: Belknap, 2001.

Rossi, E., *Modus Vivendi, Consensus, and (Realist) Liberal Legitimacy*, in «*Public Reason*», 2 (2010), no. 2, pp. 21-39.

Rossi, E., *Consensus, Compromise, Justice and Legitimacy*, in «*Critical Review of International Social and Political Philosophy*», 16 (2013), no. 4, pp. 557–572.

Rossi, E., *Legitimacy, Democracy, and Public Justification: Rawls' Political Liberalism vs Gaus' Justificatory Liberalism*, in «*Res Publica*», Online First (2013) doi:10.1007/s11158-013-9223-9.

Shklar, J., *The Liberalism of Fear*, in N. Rosenblum (ed.), *Liberalism and the Moral Life*, Cambridge, MA, Harvard University Press.

Wenar, L., *Political Liberalism: An Internal Critique*; *Ethics* 102 (1995), no. 1, pp. 32-62.