Political Liberalism: A Kantian View

Rainer Forst

Abstract: This article suggests a Kantian reading of Rawls’s *Political Liberalism*. As much as Rawls distanced himself from a presentation of his theory in terms of a comprehensive Kantian moral doctrine, we ought to read it as a noncomprehensive Kantian moral-political theory. According to the latter approach, the liberal conception of justice is compatible with a plurality of comprehensive doctrines as long as they share the independently defined and grounded essentials of that conception of justice—that is, as long as they are “reasonable,” to use the term that does most of the Kantian work.

This paper will be published in *Ethics* 128 (October 2017): 123-144.

You can find the published version for citation purposes online: https://doi-org.proxy.ub.uni-frankfurt.de/10.1086/692945

Contact: forst@em.uni-frankfurt.de
“A political conception of justice is what I call freestanding … when it is not presented as derived from, or as part of, any comprehensive doctrine. Such a conception of justice in order to be a moral conception must contain its own intrinsic normative and moral ideal.”

I. THE FAMILIAR INTERPRETATION OF POLITICAL LIBERALISM

In the following, I argue against the familiar interpretive story about Rawls's *Political Liberalism* which asserts that it marks a complete turn away from Kant. If we follow that interpretation, we will not be able to understand Rawls’s project, which is to develop a freestanding conception of justice that is justified on the basis of—noncomprehensive—principles and ideas of practical reason alone. It does not depend on any comprehensive doctrine for its validity, and it takes priority over comprehensive views since it defines autonomously whether they are reasonable or not.

To be sure, the familiar interpretation rightly stresses the differences between Rawls’s earlier Kantian theory and the later version of it. As Rawls explained, in his original theory “a moral doctrine of justice general in scope is not distinguished from a strictly political conception of justice” (*PL*, xv). Thus, from the perspective of *Political Liberalism*, Rawls would no longer regard the original position as “a procedural interpretation of Kant’s conception of autonomy and the categorical imperative within the framework of an empirical theory,” as he did in *A Theory of Justice*. Nor would he argue for the motivational “congruence” of political and social justice and individual goodness based on the Kantian idea that “acting justly is something we want to do as free and equal rational beings,” so that the “desire to act justly and the desire to express our nature as free moral persons turn out to specify what is practically speaking the same desire” (*TJ*, 501). Within such an account of our unified “practical identity,” considerations of justice guide the pursuit of all of our practical goals and thus constitute rather than simply constrain our notion of the good: “What we cannot do is express our nature by following a plan that views the sense of justice as but one desire to be weighed against others. For this sentiment reveals what the person is, and to compromise it is not to achieve for the self free reign but to give way to the contingencies and accidents of the world” (*TJ*, 503).

It is not difficult to understand why a reflection on the core aims and history of liberalism, especially when viewed in the light of the question of toleration, as Rawls increasingly did, led him to believe that this Kantian account of the theory was not only “unrealistic” (*PL*, xvi) but also contradictory. For it had to be a theory of justice for a pluralistic society, and that included, of course, a diversity of religious and philosophical doctrines. And the idea of such a society becoming a society of virtuous Kantians who struggled to liberate themselves from the contingencies of social life and regarded themselves as noumenal selves who produce values autonomously by following laws of reason seemed to go too far. The theory had to move away from the ground of comprehensive doctrines of this kind and confine itself to “political”
conceptions and values that appeared “reasonable” from and “implicit” (PL, 13) in the public culture of a democratic modern society. Political liberalism had to apply the principle of toleration to itself (PL, 10) and reduce its foundational program to one of reconstructing generally accepted conceptions of social cooperation and democratic citizenship. Or so it seemed.

At the extreme, this alleged turn away from Kant led to interpretations such as that of Rorty, who argued that Rawls’s new approach was “thoroughly historicist and antiuniversalist” and simply “a historico-sociological description of the way we live now.”⁴ In a slightly different version, we find similar thoughts expressed by Burton Dreben, who argued that Rawls gave up all attempts to ground his theory philosophically and simply tried to work out notions implicit in our tradition of democratic thought.⁵ And Dreben adds an interesting observation against which I will argue: “Kant’s talk about practical reason is useless for understanding Rawls.”⁶

II. THE PROBLEM OF POLITICAL LIBERALISM

In my opinion, this familiar interpretive story is wrong. For it overlooks that, as much as Rawls undeniably tried to distance himself from a presentation of his theory of justice in terms of a comprehensive Kantian moral doctrine, he did present it as a noncomprehensive Kantian moral-political doctrine, one which is compatible with a plurality of comprehensive doctrines as long as they share the independently defined and grounded essentials of the doctrine—that is, as long as they are “reasonable,” to use the term that does most of the Kantian work in Political Liberalism.⁷ In my view, the kinds of conventionalist, historicist, and relativistic interpretations cited cannot explain the “as long as” qualifying clause that can be found everywhere in Political Liberalism. In other words, they cannot explain the foundational role that a particular conception of practical reason plays in the book. I will explain this in what follows, though not without ultimately criticizing Rawls for not having found the adequate language to describe his new approach.

In order to explain the sense in which Rawls’s version of political liberalism is still a Kantian view, I will begin with a brief reflection on the central problem that it addresses. As already mentioned, Rawls situates his project firmly within the liberal tradition, and especially within the liberal tradition that deals with the question of toleration. Moreover, he is right to differentiate a form of Enlightenment thought that tries to replace the plurality of religions with a new, secular doctrine (or a semisecular doctrine of reasonable, “natural” religion in a deistic fashion) from other forms of Enlightenment thought that sought to leave more room for religious diversity and confined themselves to political norms that did not conflict with comprehensive doctrines with respect to metaphysical questions or questions of the ultimate values to be realized in life. In my Toleration in Conflict,⁸ I have reconstructed the discourse of toleration (in the West) and emphasized the difference between these approaches, and Rawls’s work can be seen as the most recent expression of one of these strands. In Section V, I will show how the seventeenth-century philosopher Pierre Bayle identified the same problem that Rawls addresses and developed a
proto-Kantian argument that is close to Rawls’s in that it also defends an autonomous notion of reason as the basis for a conception of justice and tolerance that binds persons with very different comprehensive views.

In explaining the problem, Rawls stresses from the outset that “political liberalism takes for granted not simply pluralism but the fact of reasonable pluralism” (PL, xviii). Thus, while he emphasizes that the political doctrine aims to be impartial between the “points of view of reasonable comprehensive doctrines” (PL, xix), he is never for a moment willing to compromise or qualify the meaning of reason or reasonableness with respect to the plurality of existing comprehensive doctrines. The notion of the reasonable is, so to speak, defined a priori or, in Rawls’s language, in a “freestanding” way, that is, prior to any comprehensive doctrine, and it autonomously qualifies which of these doctrines count as reasonable and which do not. It always takes normative priority. That is why at the beginning of the introduction to Political Liberalism, Rawls announces that the first three lectures will “set out the general philosophical background of political liberalism in practical reason” (PL, xiv). They reconstruct the Kantian program in political-moral terms that are not comprehensive. As Rawls affirms in many places in the book—and even more so in the newly added “Introduction to the Paperback Edition” three years later, in which there are abundant passages such as the one cited above as the epigraph—the political conception is, “of course, a moral conception … worked out for a specific kind of subject, namely, for political, social, and economic institutions” (PL, 11).

Thus, if the question of political liberalism is “How is it possible that there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable religious, philosophical, and moral doctrines?” then the answer is that this is possible only if the theory gets the notions of reason and of reasonableness right, and for that purpose a Kantian (and only a Kantian) would say “that the principles and ideals of the political conception are based on principles of practical reason in union with conceptions of society and person, themselves conceptions of practical reason” (PL, xx). In other words, stability “for the right reasons” (PL, xxxvii) can only be attained if the right reasons can be identified and vindicated by way of a reconstruction of practical reason. This is the program of political constructivism, and, as I will show below, even though it differs in important ways from Kant’s moral constructivism, Kantian constructivism is still its model. This is made clear by lecture III of Political Liberalism and by a comparison between that lecture and the full reconstruction of Kant in Rawls’s text on “Themes in Kant’s Moral Philosophy,” which shows that the structures are identical despite the differences.

As for any liberal view, the problem was, of course, to explain how the political conception, which provides the “reasonable public basis of justification on fundamental political questions” (PL, xix), is related to the comprehensive doctrines citizens may hold. Rawls often uses the phrase that they are “somehow related” (ibid.) to point to the problem of combining these “two views” (PL, 140)—a problem that the notion of reason has to resolve. Reason needs to take priority over comprehensive doctrines in matters of justice, but not in the same way
when it comes to metaphysical or ethical matters of the good, and thus both justice and a comprehensive doctrine need to be affirmed by persons sincerely and “from the inside,” so to speak. Reason has to constrain ethical comprehensive doctrines morally when it comes to matters of justice, but at the same time it must not colonize them when it comes to the issues that only comprehensive doctrines can answer. When we turn to Bayle, we will see how he tried to solve this problem—and how close his solution is to Rawls’s view.

III. KANTIAN CONSTRUCTIVISM IN POLITICAL LIBERALISM

Let us take a closer look at the Kantian character of Rawls’s enterprise. A first, major issue is the “freestanding” character of the political (and this generally means: moral-political) conception of justice. That a theory of justice ought to be independent with respect to philosophical debates in metaphysics, for example, was a long-standing thesis of Rawls’s dating back to the seventies.10 In the context of political liberalism this becomes a special, far-reaching kind of independence. Yet there are two ways to look at this. The first, meliorating (or accommodating) perspective is that a freestanding conception of justice that “is neither presented as, nor as derived from” (PL, 12) a comprehensive doctrine is not in conflict with such a doctrine, since it avoids rival implications with regard to the truth regime of the doctrine (such as the existence of God, of evil, and so on). So it is on a peaceful footing with the doctrine and can fit into it as “a module, an essential constituent part” (ibid.). But the other perspective—let us call it the priority perspective—is no less, and perhaps even more, relevant. For only an independently grounded conception of norms of justice, that is, one that is grounded on reason, can generate the normative force to determine which of the comprehensive doctrines is reasonable and which is not; it is the umpire on these questions, the only authority there can be. This is the Kantian aspect, and it is remarkable how many interpreters overlook or play down this aspect—or see it as an aberration.11 It is the reason why Rawls can rescue his theory from the charge that it is “political in the wrong way” (PL, 40), that it is a mere modus vivendi or compromise between comprehensive doctrines, and claim instead that it is a “freestanding view … working from the fundamental idea of society as a fair system of cooperation and its companion ideas.” And Rawls adds an important statement: “We leave aside comprehensive doctrines that now exist, or that have existed, or that might exist” (ibid.). The priority view is essentially a Kantian view, following Kant in emphasizing that both the categorical imperative and the principle of right had to be grounded completely independently of any doctrine of value leading to the good life (or Glückseligkeit) in order to take priority over them.

This aspect of the approach culminates in the thesis of “doctrinal autonomy” (PL, 98) which Rawls states in expounding constructivism in lecture III, where he connects the notion of “freestanding” with that of autonomy in a Kantian way. He argues that the constructivist program “enables us to state the meaning of an autonomous political doctrine as one that represents, or displays, the political principles of justice … as reached by using the principles of practical reason in union with the appropriate conceptions of persons as free and equal and of
society as a fair system of cooperation over time. ... Think of this as doctrinal autonomy” (ibid.). And in a fashion characteristic of his noncomprehensive, political Kantianism, he adds that the view is autonomous because it is based only on the practical reason of citizens used in grounding and understanding a political conception of justice, and thus “in affirming the political doctrine as a whole we, as citizens, are ourselves autonomous, politically speaking” (ibid.). This is the core of what Rawls calls “full autonomy” in a political sense—not an ethical conception that guides persons in their personal life choices and considerations of the good life, but nevertheless a moral one, because it expresses essential moral duties of citizens when it comes to fundamental questions of justice within the basic structure of the society to which they belong.

The program of an autonomous grounding of a theory of justice is laid out in the lecture on constructivism. The main idea of constructivism is to establish a procedure of construction based on practical reason to generate justifiable norms (see PL, 90) that no reasonable person can deny. Rawls makes a plausible distinction between a constructivist comprehensive moral doctrine (especially Kant’s own) and political constructivism, on the grounds that the latter (a) is not committed to the metaphysical idea that such construction produces an order of values that did not exist before (instantiating what Rawls calls “constitutive autonomy,” as in Kant; see PL, 99) and (b) does not extend to “all of life” (ibid.), but only to political questions of justice within a basic structure. Still, while this characterization excludes metaphysical and ethical constitutivist constructivism, it does not sufficiently stress that the political construction is still a moral one, even though this is Rawls’s view.12 He failed to distinguish, as he should have, between two notions of the moral, one connected to a comprehensive doctrine (which I would prefer to call “ethical”) and one connected to the grounds and normative quality of the political conception—though it is obvious that he used “moral” in both of these quite different senses.

Rawls leaves no doubt that the autonomous construction of a political conception of justice relies on a procedure—the original position—that “embodies all the relevant requirements of practical reason and shows how the principles of justice follow from the principles of practical reason in union with conceptions of society and person, themselves ideas of practical reason” (PL, 90). Like Kant,13 Rawls believes that the procedure itself is not constructed, but is instead reconstructed or “assembled” (PL, 108) from reflection on our “powers of reason” (PL, 96) and is “laid out” (PL, 103) in designing the original position—or, in Kant’s case, the categorical imperative as a procedure—its basis being “the conception of free and equal persons as reasonable and rational, a conception that is mirrored in the procedure.”14 In explaining political constructivism, Rawls uses exactly the same words as in the original Dewey Lectures, though now redefined with respect to the aim of justifying principles of justice for the basic structure of a pluralistic society, thus relying on “the fundamental idea of a well-ordered society as a fair system of cooperation between reasonable and rational citizens regarded as free and equal” (PL, 103). Rawls distinguishes ideas of reason, such as that of society as a fair system of cooperation and that of the person with the two moral powers (explained in lectures I and II),
and principles of practical reason as principles of using reason as well as rationality, which the original position models in a particular way by imposing reasonable constraints on the rational choice of the parties. The main notion of the reasonable is that of reciprocity of justification, which implies that the reasons for organizing the basic structure in a particular way have to be justifiable between free and equal citizens notwithstanding their comprehensive doctrines, solely on the basis of their common practical, public reason. This also becomes the core of the notions of legitimacy and public reason spelled out in the later lectures in the book.\(^{15}\)

The Kantian character of this approach—despite the reduction to the political—is apparent in many ways, not just if one compares Rawls’s analysis of Kant’s constructivism in detail with his own approach,\(^{16}\) where many structural parallels and common thoughts can be found almost verbatim, though I cannot offer such a detailed comparison here. An important point of commonality despite important differences is to be found where Rawls rejects Kant’s notion of constitutive autonomy but accepts his view that “the principles of practical reason originate … in our moral consciousness as informed by practical reason. They derive from nowhere else” (\(PL, 100\)). Again, the independence of the political conception from comprehensive doctrines does not weaken but instead strengthens its foundations because reason is autonomous (“self-originating and self-authenticating”; ibid.) and does not need any other normative source to bind moral persons—categorically, we may add, because no other comprehensive system of value can justifiably trump the normativity of reason and its constructions. What often sounds like a modest reduction of justificatory claims for the political conception now appears as what it is: the autonomous rule of reason for autonomous persons in the realm of political and social justice.

IV. THE IMPOSSIBILITY OF A PRACTICE-DEPENDENT HERMENEUTICS

One may wonder how the Kantian character of the theory is compatible with the Rawlsian claim that the political conception relies on “fundamental ideas seen as implicit in the public political culture of a democratic society” (\(PL, 13\)), a claim which the conventionalist or historicist interpretations usually cite. But in no way is a conventionalist program of justification lurking here. For Rawls never says that these fundamental ideas are in fact guiding current practice or are widely shared in contemporary democratic societies, nor does he say that the theory of justice uses them because they are generally shared or factually present. All that Rawls says is that these are ideas implicit in a democratic society if this society can justifiably claim to be democratic at all, and there are a number of passages where—as in his explanation of the original program of reflective equilibrium—he points out that some hard work of abstraction is required to arrive at these ideas, which are presented as ideas of practical reason, reconstructed reflexively and neither arrived at by way of a metaphysical notion of rationality nor interpreted in a conventionalist mode.

When Rawls asks how a shared basis for justifying a conception of justice can be found, he suggests that we “collect” (\(PL, 8\)) historical and present beliefs conducive to justice, such as the
rejection of slavery or religious toleration, as “provisional fixed points that any reasonable conception must account for” (ibid.). The phrase “account for” suggests that the order of justification is not from historical facts or beliefs to the justification of the theory but rather the other way around: the theory has to provide independent normative reasons for such progressive and emancipatory ideas and fit them into a general account of justice. This is the task of reflective equilibrium (ibid.). Normally, Rawls states, we find public political culture “of two minds at a very deep level” (PL, 9), and thus the theory has to be based on fundamental ideas that solve such conflicts with an eye to what justice reasonably demands. There can be no purely hermeneutic or “practice-dependent” grounding of principles of justice, especially not if the aim is to express the “shared and public political reason” (PL, 9) of citizens, because there is no such shared reason as a fact, which historicist readings falsely assume. And even if there were a strong consensus, that fact would not provide sufficient reason to call the consensus reasonable, since reason must be capable of criticizing any factual normative consensus. Any candidate for such consensus needs to be argued for with reasons which are independent of existing comprehensive doctrines and, we may add, of ideological delusions that deny the freedom and equality of persons and are not the result of free public reason guided by reciprocity. Hence, the aim of the construction is to find ideas and principles that democratic citizens “can endorse” (PL, 10) if they reason properly. The ideas and conceptions of reason are shareable among reasonable persons; they are not called reasonable because many people actually share them. This is where reason acquires its critical force, a force that can be directed against ideologies: “The criterion of reciprocity requires that when those terms are proposed as the most reasonable terms of fair cooperation, those proposing them must also think it at least reasonable for others to accept them, as free and equal citizens, and not as dominated or manipulated, or under the pressure of an inferior social position.”

Rawls asserts that in “political philosophy the work of abstraction is set in motion by deep political conflicts” (PL, 44) and that “we turn to political philosophy when our shared political understandings, as Walzer might say, break down, and equally when we are torn within ourselves” (ibid.). As an example of the reflection he has in mind—and that is, of course, very different from Walzer’s own, as Rawls makes clear—he cites “Alexander Stephens rejecting Lincoln’s appeal to the abstraction of natural right and replying to him by saying: the North must respect the South’s shared political understandings on the slavery question. Surely the reply to this will lead into political philosophy” (PL, 45). This is as antihistoricist a statement as you can make, but at the same time it is grounded in a proper historical reflection on the possible ideological use of conventionalism and the need for radical moral argument at a time when the conviction about the abolition of slavery was not a shared understanding. And even if it had been, we would still need to know why this was right apart from the fact that it was a shared belief.

Rawls goes on to remind us that “political philosophy cannot coerce our considered judgments any more than the principles of logic can” (ibid.), which, if you think about it, is not such a weak claim to make about what political philosophy can do. But nowhere did Rawls claim that our
societies are already based on ideas leading to a well-ordered society or are, as Rorty thought, to a
large extent well ordered. His exercise was a philosophical one, not a hermeneutic one of
bringing out what everyone already knows and thinks. The philosophical exercise is aware that
our societies are in deep conflict about justice and that philosophy, insofar as it reconstructs
progressive ideas that are implicit in a democratic culture, needs to explain what it would mean
to regard society “as a fair system of cooperation over time. Seen in this context, formulating
idealized, which is to say abstract, conceptions of society and person connected with those
fundamental ideas is essential to finding a reasonable political conception of justice” (PL, 46).

V. TOLERATION AND REASON

It is an important step of critical self-reflection for a liberal doctrine that aims to be compatible
with a plurality of notions of the good to consider whether this is indeed the case or whether
the theory is based on a doctrine that would exclude religious or metaphysical views which a
liberal society should include. So that question for liberalism is far from “new,” as Dreben thinks,
nor is it true that it had “never been said before in the history of philosophy” that only
oppressive power can force a society to unite on one comprehensive doctrine. The history of
toleration—not just of liberal views on toleration—abounds with such reflections.

The project of Political Liberalism begins with two questions which the book unites and seeks to
answer. The first is the familiar one which asks what is the best conception of justice “for
specifying the fair terms of cooperation between citizens regarded as free and equal” (PL, 47).
The second asks what are the “grounds of toleration” for a society marked by the “fact of
reasonable pluralism” (ibid.). So the second question is not about the possibility of just any kind
of toleration, including modus vivendi arrangements, but about a “ground” of toleration that is
strong enough to support a freestanding conception of justice but is nevertheless compatible
with a plurality of reasonable doctrines. Thus, it applies the principle of toleration to “philosophy
itself” (PL, 10): a system of norms for social cooperation has to be fair and acceptable to all
those comprehensive doctrines that respect all citizens as free and equal and seek productive
social cooperation. The conception of justice has to be acceptable to all reasonable doctrines
and define independently what “reasonable” means in noncomprehensive terms. Otherwise, the
doctrines would quarrel endlessly about the core principles and implications of justice, and no
public form of reason could ever exist.

In order to understand how much this question aims at the heart of a combination of justice
and toleration that has been the focus of centuries of historical struggles, allow me a historical
argument which helps to arrive at the correct interpretation of the complex notion of
reasonableness in Political Liberalism. Rawls was a great interpreter of the history of philosophy,
but I don’t think he ever came across Pierre Bayle in his research. He found other kindred spirits
like the Bodin of the Colloquium of the Seven, where Bodin was one of the first to show that there
can be a highly reasonable debate between very different religions and metaphysical views
without one showing the others to be unreasonable or clearly wrong judging on the basis of
general laws of reason—a good example for what Rawls calls “reasonable disagreement.”

But it is Pierre Bayle (1647–1706) who presents a theory of toleration that is both proto-Kantian
and proto-Rawlsian in important (though not all) respects relevant for our discussion. Bayle,
as a Huguenot who suffered persecution in France and an undogmatic thinker who later also
became a target of the hostility of his fellow believers, recognized that a justification of
toleration had to include a reciprocal duty of tolerance that needed to be morally justified
independently of specific articles of religious faith. Otherwise, the endless strife over who was in
the right and was allowed to coerce others in the name of the “true religion” could never be
resolved. At the same time, however, such a justification should not come at the expense of the
convictions of each party to the dispute that they were advocating the true faith; religious
skepticism was not a viable solution. Still, it had to be possible to arrive at the shareable moral
insight that it is “childish” always to insist only on one’s own truth and the authority to suppress
others in social conflicts, since that truth is precisely what is in dispute. There had to be a form
of practical reason which made it clear that, without independent and shared principles, any act
of violence could be deemed godly.

Bayle develops these ideas in detail in his *Commentaire philosophique* (1685). There he argues that
the “natural light” of reason, which God has implanted in all human beings independently of
their religion, reveals the “most general and infallible principles” of morality: “But since passions
and prejudices only too often obscure the ideas of natural equity, I would advise a person who
intends to know them well to consider these ideas in general and as abstracted from all private
interest and from the customs of his country.” One should then ask oneself whether a certain
practice could meet with universal agreement: “Is such a practice just in itself? If it were a
question of introducing it in a country where it would not be in use and where he would be free
to take it up or not, would one see, upon examining it impartially that it is reasonable enough to
merit being adopted?”

Bayle argued that proponents of forced conversions and persecution were inverting the
requirements of morality and turned virtues into vices. The mistake was the presumption that
one has the right to impose the true religion by force, so that violence suddenly becomes “good”
or “salutary.” According to Bayle, this is “the most abominable doctrine that has ever been
imagined.” With this argument, anyone could turn any position on its head:

If one would say, “it is very true, Jesus Christ has commanded His Disciples to persecute, but
that is none of your business, you who are heretics. Executing this commandment belongs only
to us who are the true Church,” they would answer that they are agreed on the principle but not
in the application and that they alone have the right to persecute since truth is on their side. …
When one reflects on all this impartially, one is reduced necessarily to this rare principle, I have
truth on my side, therefore my violences are good works. So and so errs: therefore his violences
are criminal. To what purpose, pray, are all these reasonings? Do they heal the evils which
persecutors commit, or are they capable of making them reconsider? Is it not absolutely necessary in order to cure the furor of a zealot who ravages a whole country or to make him comprehend his doings, to draw him out of his particular controversies and remind him of principles which are common to both parties such as the maxims of morality, the precepts of the Decalogue, of Jesus Christ and of His Apostles, concerning justice, charity, abstinence from theft, murder, injuries to our neighbour, etc.\textsuperscript{230}

There are two key components of Bayle’s argument for toleration: the \textit{normative} component of the morality of reciprocity and the \textit{epistemological} component of the nondemonstrability of the undeniably true faith by means of reason alone. For violence on “natural” moral concepts remains mere violence, and the claim to speak for the unquestionably true religion cannot be redeemed by “natural” reason on grounds that cannot be reasonably rejected. According to Bayle, it is not just a matter of appealing to an independent, rational sense of morality that is free from fanatical distortions and is shared by all human beings, in order to be able to differentiate moral from religious truths. It is also a matter of undercutting religious disputes by showing that, although they are not pointless, they cannot be resolved here on earth by rational means alone. This calls for a conception of the \textit{finitude of reason} which states that disagreements among finite rational beings in questions of faith are unavoidable.

Bayle defends a conception of finite practical and theoretical reason whose guiding assumption is that reason must recognize its own limits regarding “speculative truths.” This opens up the space of metaphysical or religious conflict between positions that can be \textit{reasonably held} but can also be \textit{reasonably rejected}. The reason is that “evidence is a relative quality” especially in religious matters.\textsuperscript{31} Habit, training, or other factors mean that rational individuals arrive at very different evaluations and judgments. A reasonable person is aware, we might say here, of the “burdens of reason” (or “judgment”), to use Rawls’s phrase, and, according to Bayle, knows that “difference in opinion [is] man’s inherent infelicity, as long as his understanding is so limited and his heart so inordinate.”\textsuperscript{32} Therefore, the desire that all human beings should unite in one religion will remain unfulfilled, and the reasonable response is to espouse toleration. Rational human beings recognize that their reason is finite and that religious differences are rationally unresolvable.

This is the central theme of Bayle’s \textit{Dictionnaire historique et critique} (1696).\textsuperscript{33} His main concern in this work is to create room for religious answers to metaphysical questions by placing limits on the force of reason. This cuts the ground out from under dogmatic disputes about and alleged proofs of the “true faith,” without faith, which remains within the boundaries of what can be rationally debated, becoming empty or irrational as a result. Both sides, reason and faith, must heed their respective limits: reason recognizes its limitations in speculative matters to which faith alone can provide further answers, and faith does not try to present and impose its “truths” as conclusive matters that are beyond reasonable dispute. Reasonable faith knows that it is a faith; it is aware that “the mysteries of the Gospels are above reason \textit{dessus de la Raison}.” That is an
insight, Bayle continues, into the “limits” of reason which “can never attain to what is above it.”

VI. RELATING THE POLITICAL CONCEPTION AND COMPREHENSIVE DOCTRINES IN THE RIGHT WAY

If we take the two main—normative and epistemological—aspects of Bayle’s theory of toleration based on reason into account, it becomes apparent how they help us understand the two aspects of the reasonable in Rawls’s account in lecture II of *Political Liberalism*. Like Bayle, Rawls distinguishes between a normative and an epistemological aspect of the reasonable, with the first being accorded priority (given the priority of practical reason). The first aspect says that persons are reasonable “when, among equals say, they are ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so” (*PL*, 49). The second aspect is “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” (*PL*, 54). These are precisely the two aspects of reason that Bayle thought were necessary to establish, in Rawlsian language, a “public and shared basis of justification” (*PL*, 61) for the normative basic structure of a pluralistic society and to explain how reasonable citizens can respect each other as reasonable and politically autonomous and cooperative agents, even though they differ deeply in their comprehensive doctrines. Insofar as we are reasonable, we recognize in particular “that our own [comprehensive] doctrine has, and can have, for people generally, no special claims on them beyond their own view of its merits. Others who affirm doctrines different from ours are, we grant, reasonable also, and certainly not unreasonable” (*PL*, 60).

This is the same lesson that Bayle’s reflection on the difference between reason and faith was intended to teach, and both Bayle and Rawls extend this to other metaphysical disputes about, say, the sources of evil or the ultimate meaning of life. Reason has a stake in these debates, insofar as it tries to draw the line between reasonable forms of faith and unreasonable—that is, immoral and irrational (superstitious)—ones, but by its own powers it cannot resolve these debates; it can neither prove nor finally reject any of the reasonable doctrines. But reason remains aware of its own powers when it comes to matters of morality—and in Rawls’s case, we should say, to matters of political morality, which is one of the differences from Bayle, who had a more expansive conception of the realm of morality. So my point is not to deny that Bayle, both in this expansion and in his own seventeenth-century rationalistic view of the sources of reason, was also defending a “comprehensive” doctrine in some way. My point is that, in an early form, he saw exactly the problems of grounding a scheme of justice and toleration on a doctrine that rivals with religious ones on the same level.

It is striking how close Rawls’s explanation of the “burdens of judgment” is to Bayle’s explanation of the fact of a reasonable plurality of religious and metaphysical views by
reminding us of relative evidence, habit, different upbringing, and so on. Rawls lists exactly the same difficulties in assessing evidence, difference in evaluations, and indeterminacy in hard cases, the influence of biographical experience and socialization on our judgments, and so on, when he lists the limits “to the theoretical uses of our reason” (PL, 56). Reflection on the limits of reason in metaphysical and religious matters already makes one aware that “reasonable disagreement” is a normal condition of life, especially in a pluralistic society, but reciprocity of justification does not automatically follow from this. For this, the first aspect of practical reason is essential. Here are the Baylean conclusions drawn by Rawls:

Those who insist, when fundamental political questions are at stake, on what they take as true but others do not, seem to others simply to insist on their own beliefs when they have the political power to do so. Of course, those who do insist on their beliefs also insist that their beliefs alone are true: they impose their beliefs because, they say, their beliefs are true and not because they are their beliefs. But this is a claim all equally could make; it is also a claim that cannot be made good by anyone to citizens generally. … It is unreasonable for us to use political power, should we possess it, … to repress comprehensive views that are not unreasonable. (PL, 61)

Rawls has often been criticized for the seeming “schizophrenia” of his idea that citizens regard other views they find wrong (insofar as they consider the comprehensive doctrine they themselves hold to be true) nevertheless to be reasonable in practical and theoretical terms. Others have described this view—and the reduction of the truth claims of the political conception—as one of “epistemic abstinence,” which mistakes political philosophy for some kind of accommodating politics. But there is nothing contradictory or empty here; Rawls simply explains, as any liberal should, how a devout Catholic can respect a Muslim as an equal and reasonable citizen while still rejecting Islamic faith as false. He shows precisely the possibility of what Brian Barry doubted, namely, that “certainty from the inside about some view can coherently be combined with the line that it is reasonable for others to reject that same view.”

As Bayle and Rawls argue, the kind of toleration that goes along with this is not doomed to skepticism, as you retain your belief in the truth of your doctrine if you respect others as not unreasonable. But if it is a religious doctrine and you think that it is demonstrably the only one reasonable persons can hold with good, rational reasons, you are being dogmatic and have not understood the nature of religious faith and disagreement as viewed from a reasonable perspective. Rawls provides an insight into the core of the connection between reason, justice, and toleration, and it in no way diminishes either the truth claims of religions (as some worry) or the independent grounding of his own conception (as others worry). But this becomes apparent only if one understands the notion of the reasonable in all of its facets and strengths. This is what Bayle and Rawls share with Kant, whose philosophy brings out the autonomy of reason in all its clarity.
Arguing that there is nothing schizophrenic about relating the “two views” (PL, 140) of citizens, as Rawls expresses it, does not mean that there is an easy explanation of how this is achieved within the perspective of one and the same person. Rawls addresses this issue in terms of the notion of an “overlapping consensus,” which plays an important role in explaining the stability of a well-ordered pluralistic society after the foundations in practical reason have been laid out. But the account of stability generated by an overlapping consensus has often been misunderstood, as though the only grounds for accepting the political conception were those from within the comprehensive doctrine, leaving no shared moral substance to the conception of justice. For from the start we speak only of a “reasonable overlapping consensus” (PL, xlvii) which neither is a modus vivendi compromise nor comes about by “striking a balance” between existing comprehensive doctrines; rather, according to Rawls, “we formulate a freestanding political conception having its own intrinsic (moral) political ideal expressed by the criterion of reciprocity” (ibid.). Rawls considers the “values of the special domain of the political” as a “subdomain of the realm of all values” (PL, 139) and leaves the latter to be comprehensively determined by the various ethical doctrines, but as reasonable doctrines they all accept that the political values “normally outweigh whatever values may conflict with them” (ibid.). This is because the “values of the political are very great values and hence not easily overridden” (ibid.): they constitute the shared framework of political and social life for citizens, and reasonable citizens know that they owe each other the duty to establish and preserve justice when it comes to this framework. This is what their sense of justice, as an essential characteristic of citizens, tells them. So it is not correct to say that the overlapping consensus is the only answer to the question of stability that Rawls addresses in the second part of Political Liberalism, for in this context he also reminds us of the independent normative quality of the “ideal of citizenship” (PL, 84) and its motivating force: “Citizens in a well-ordered society acquire a normally sufficient sense of justice so that they comply with its just arrangements” (PL, 141). Hence, just as the “principles of justice are not affected in any way by the particular comprehensive doctrines that may exist in society” (ibid.), so too citizens are autonomously motivated to accept these principles. This is the meaning of “full autonomy” discussed above.

Thus, when Rawls explains the different comprehensive grounds—some religious, some based on moral doctrines—and shows how reasonable citizens affirm the political conception from within their own view, he is not saying that these are the only grounds on which they do so. For, insofar as they are reasonable, they always recognize their duties of justice, and they accept the political conception as a “moral conception” (PL, 147). In a crucial passage, where Rawls argues (again) against the modus vivendi interpretation, he adds that the political conception as a moral conception “is affirmed on moral grounds, that is, it includes conceptions of society and of citizens as persons, as well as principles of justice, and an account of the political virtues through which those principles are embodied in human character and expressed in public life” (ibid.). It is with these grounds in place that citizens also draw on their comprehensive doctrines when they affirm the political conception, and Rawls adds that this “does not make their affirming it any
less religious, philosophical, or moral” (PL, 147–48), which follows from a successful integration of the political conception into a comprehensive doctrine as a “module, an essential constituent part” (PL, 12) of that doctrine. But that does not mean that by doing so persons are no longer aware of the political values and the moral character of the conception as binding on reasonable citizens generally: there is no gestalt shift in the reasons from the moral-political conception to the comprehensive doctrine; rather, the comprehensive doctrine shows its reasonableness by integrating and at the same time preserving the binding force of the conception of justice.

This is why, as Rawls goes on to say, “those who affirm the various views supporting the political conception will not withdraw their support from it should the relative strength of their view in society increase and eventually become dominant” (PL, 148). Hence, they will accord justice priority over the good even if their religion tells them something different, since they know that it would be unreasonable and unjust to do otherwise. That priority could not be upheld and affirmed if the political conception was no longer an independent moral force. Another passage makes this clear: “Thus the political conception can be seen as part of a comprehensive doctrine but it is not a consequence of that doctrine’s nonpolitical values. Nevertheless, its political values normally outweigh whatever other values oppose them, at least under the reasonably favorable conditions that make a constitutional democracy possible” (PL, 155; emphasis added). So we must not misunderstand what it means to “apply the principles of toleration to philosophy itself” (PL, 154): it means to seek moral grounds for a political conception of justice that citizens cannot reasonably reject (to use Scanlon’s phrase) and that leave room for all of the religious or ethical or metaphysical answers to questions that point beyond the realm of the reasonably nonrejectable without thereby being unreasonable.43 “Thus, the values that conflict with the political conception of justice and its sustaining virtues may be normally outweighed because they come into conflict with the very conditions that make fair social cooperation possible on a footing of mutual respect” (PL, 157). Only a Kantian reading, I believe, can make sense of the meaning and priority of “mutual respect” mentioned here.

In his “Reply to Habermas,” Rawls also asserts that there is an independent pro tanto justification of the political conception of justice “without looking to, or trying to fit, or even knowing what are, the existing comprehensive doctrines.”45 In “full justification,” that conception gets “embedded” into the comprehensive doctrines of persons individually, and that responds to the task of relating political and nonpolitical values in the right way, for which, Rawls affirms, the political conception gives no ethical or comprehensive guidance. Reason, however, which defines a reasonable comprehensive doctrine, does provide such guidance: it integrates the political conception and the other comprehensive aspects of the doctrine in the proper way, because the doctrine is reasonable not just from the perspective of an outside observer’s description but also, so to speak, “from the inside,” as a personal-political, reflexive point of view. This is why “public justification” can take place on a third level of justification, where citizens debate issues of justice and where “the shared political conception is the common ground.”46 And indeed we should ask, how would that kind of public justification, or what
Rawls calls the exercise of “public reason,” be possible if the political conception did not serve as such a common ground, constraining other aspects of the comprehensive doctrines?

Rawls’s view of public reason changed over time as regards the question of how permissive it could be of reasons that stem from comprehensive doctrines and that are not based on political values of the political conception alone (when it comes to essential questions of justice). But, be that as it may, the whole approach presupposes an “ideal of citizenship” (PL, 213) where citizens give strict priority to political values and reasons and accept—as the liberal principle of legitimacy asserts—that “our exercise of political power is proper and hence justifiable only when it is exercised in accordance with a constitution the essentials of which all citizens may reasonably be expected to endorse in the light of principles and ideals acceptable to them as reasonable and rational” (PL, 217). Here we see that the justification of the political conception in the constructivist procedure also grounds the very possibility of political legitimacy and justification. For if citizens were completely caught up in their comprehensive doctrines as their only view, the exercise of public justification would be a burden they could not shoulder; in fact, they would lack the perspective on which it relies, the perspective of a shared conception of justice. So they accept the “duty of civility” to act in accordance with the principle of legitimacy as “a moral, not a legal, duty” (ibid.). And even in the “wide” view of public reason that Rawls later espoused, citizens have the duty and ability to distinguish between comprehensive and public reasoning, and comprehensive views may only be introduced in public reasoning “provided that in due course public reasons, given by a reasonable political conception, are presented sufficient to support whatever the comprehensive doctrines are introduced to support” (PL, xlix–l). Again, it is not conceivable that persons could be held to this duty if they did not have an independent and effective understanding and sense of justice based on practical reason—a capacity all citizens are seen as sharing. Otherwise, as Bayle, Kant, and Rawls fear, a public conception of justice would not develop; rather, questions of justice would be the subject of constant conflicts between comprehensive doctrines. Justice may be, to use a metaphor, a diamond that shines in different colors depending on the plurality of comprehensive views directed at it, but its intrinsic worth does not depend on the light shone on it by the comprehensive views.

VII. AMBIGUITIES

In conclusion, I would like—making a long story very short—to remark on a fundamental ambiguity in Rawls’s theory I already hinted at. I hope to have shown that Political Liberalism is best read as a Kantian view, that is, as one which conceptualizes a noncomprehensive, autonomous, morally grounded theory of political and social justice for a pluralistic society. It is noncomprehensive in that it neither rests on some metaphysical notion of human nature nor seeks to give guidance on questions of the good life. It is autonomous in that it is based on practical reason as the capacity of autonomous citizens who respect each other as free and equal
to reciprocally and generally justify and accept principles of justice. And it is moral insofar as it has an independent normative force that is strong enough to outweigh other, competing values.

The problem, however, is that Rawls did not fully develop the conceptual tools he required for this project. Most importantly, he did not distinguish between Kantian constructivism in moral theory all the way down (e.g., as denying moral realism) and Kantian constructivism in political theory as a matter of avoiding metaphysical claims, as well as norms governing life as a whole. But above all, Rawls did not find a terminology for distinguishing between the notion of morality he required and used for the political conception and the notion of morality that was part of a comprehensive view. One could follow Habermas and others like Dworkin or Williams and use “moral”—or better, “moral-political”—for the first and “ethical” for the second, or phrase this in some other way. There are different ways to conceptualize this terminologically, but to avoid ambiguities a distinction needs to be made. For as much as Rawls stressed that the political conception is a “moral” conception that contains “its own intrinsic normative and moral ideal” (PL, xlv), he also emphasized that the political conception is affirmed on “moral” grounds stemming from the comprehensive doctrines of persons (see PL, 148). Of course, both can be the case, depending on how one looks at it, since from a personal perspective a conception of justice can be affirmed on more than one ground as long as there is the required “common ground” between citizens; however, by not making the difference more explicit, Rawls did not distinguish clearly between these different meanings of the word “moral.” Thus, at times his Kantianism is more veiled than outspoken.

Still, the justification, design, and implications of the theory are best explained by regarding it as having a noncomprehensive Kantian character. One may think, as I do, that such an approach is the most promising one for theorizing political and social justice, and one may think, as I also do, that Rawls’s approach exhibits a number of problems we ought to avoid. But he was right that any such approach needs to apply the principle of toleration to itself in the right way, responding to (reasonable) ethical pluralism while still holding on to a moral-political conception of justice based on practical reason. For what better ground can a theory of justice have than that one?

NOTES

*. I am grateful to the participants of the Rawls Conference at Georgia State University in May 2016 for an extremely helpful discussion of the version of this article presented there—with special thanks to Andrew Altman for his commentary. I am also indebted to Henry S. Richardson, Lea Ypi, and two reviewers for Ethics for their excellent comments and suggestions.


7. Note that my interpretation does not say that Rawls uses “Kant’s ideas of practical reason” in *Political Liberalism*, which Rawls calls a “serious mistake” in his letter to his editor from 1998, reprinted in *PL*, 438. In the same letter, he affirms the view held in the essay “The Idea of Public Reason Revisited,” which exhibits exactly the Kantian structure of argument I highlight. I discuss this essay below.


11. For the latter, see Wenar, “Political Liberalism: An Internal Critique.”


14. Ibid.

15. In the “Introduction to the Paperback Edition,” Rawls states that the “criterion of reciprocity” says that “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 a justification of those actions” (PL, xlvii) and adds that the duty to follow that criterion “is a duty arising from the idea of reasonableness of persons” (ibid., n. 14). In my own work, I explain the criterion of reciprocity in justification in a different but related way, linking it to a categorical duty of justification; see Rainer Forst, *Justification and Critique: Towards a Critical Theory of Politics*, trans. C. Cronin (Cambridge: Polity, 2014), chaps. 1 and 4, and *Right to Justification*, pt. 1.


19. See also John Rawls, “Reply to Habermas,” *Journal of Philosophy* 92 (1995): 132–80, 138: “No sensible view can possibly get by without the reasonable and rational as I use them. If this argument involves Plato’s and Kant’s view of reason, so does the simplest bit of logic and mathematics.”

20. The emphasis on the importance of abstraction is also the reason why I disagree with O’Neill’s claim that the constructivist approach in *Political Liberalism* is “internal to a bounded society … rather than universal or cosmopolitan,” i.e., “more Rousseauian than Kantian” (O’Neill, “Constructivism in Rawls and Kant,” 353).


22. See Forst, *Toleration in Conflict*, where I reconstruct such reflections throughout Western history.


24. One important difference is that Bayle did not develop a liberal political theory, as he thought that only a strong sovereign like Henri IV could provide a safe framework for social toleration.


27. Ibid., 31.
28. Ibid., 30.
29. Ibid., 47.
30. Ibid., 84–85.
31. Ibid., 93.
32. Ibid., 141.
34. Ibid., 410–11.
35. This is why Rawls says that being reasonable “is not an epistemological idea (though it has epistemological elements)” (PL, 62).
36. It is interesting that Rawls, directly after the remark about claims to religious dominance and before the (thoroughly Baylean) reciprocity counterargument, inserts a footnote citing Bishop Bossuet defending Catholic persecution by appealing to religious truth after the Edict of Nantes had been revoked in 1685—the same Bossuet whom Bayle attacked in his Commentaire written in the same year, a famous speech by the bishop having provided Bayle with the reason to begin the book.
40. The burdens of judgment do not deny, as Wenar believes, that religious truth is “accessible to all clear minds and open hearts” (“Political Liberalism: An Internal Critique,” 46). They only imply that religious faith requires “a religious assent of the soul” (ibid., 45, citing the Catholic teaching of Lumen gentium from 1964) that reason alone cannot produce, as it requires belief in divine revelation.
41. I won’t discuss Kant’s view of toleration at this point, but see Forst, Toleration in Conflict, sec. 21.
42. Among many others, see Weithman, Why Political Liberalism?, 308.
43. I present my own theory of toleration as a “tolerant” one in Forst, Toleration in Conflict, pt. 2.
44. I discuss the debate between the two in detail in Forst, Right to Justification, chap. 3.

46. Ibid., 144.

47. I discuss this in Forst, Contexts of Justice, chap. 3.1.


50. I discuss these in Forst, Contexts of Justice, chaps. 3 and 4, as well as in Right to Justification, chap. 3.