Some Background

It is not often in the history of philosophy that two great contemporary living philosophers engage one another in dispute. Sidney Morgenbesser, Special Projects editor of the *Journal of Philosophy*, would have known that. When he found out about the conference on Rawls taking place in Bad Homburg in the Summer of 1992, at which Habermas was to participate, he wrote to Rawls proposing a special issue of the journal. The idea was that Rawls and Habermas would “publish articles on each other’s work as it relates to questions of justice” (Letter to Rawls, 31 July 1991). Rawls generally avoided getting involved in such disputes, finding them unfruitful, but embraced this opportunity and replied that he would talk to Habermas personally about the format. He did. The dispute envisaged by Morgenbesser did not eventuate. Habermas was too busy completing his own book, *Faktizität und Geltung*, and chose to limit his contribution to the conference to the posing of some critical questions to Rawls. The eventual format of the exchange, which finally appeared in the *Journal of Philosophy* in 1995, was far more modest: Habermas offered some critical comments on Rawls’s *Political Liberalism*, and Rawls replied.

That the 1995 exchange took the form of a critical review and a reply is obvious from the titles of their respective contributions: Habermas’s “Reconciliation through the Public Use of Reason: Remarks on John Rawls’s Political Liberalism” and Rawls’s “Political Liberalism: Reply to Habermas.” Yet it is surprising how many people don’t notice this, and don’t take into account that the eventual exchange was nothing like the dispute that Morgenbesser originally envisaged, the dispute over issues of common concern between Habermas’s and Rawls’s mature political theories. I think this has something to do with the general sense of disappointment expressed by political theorists in the aftermath of the exchange. The “Habermas-Rawls accord” was, claimed Andrew Kuper, “largely a failure.” Onora O’Neill concurred, noting that this was because there are “too many premises heading in different directions.” Jonathan Wolff spoke of “the failure of the two greatest minds to meet” (Wolff 2008). Joseph Heath, a sympathetic scholar of Habermas, talks of a “relatively low level of philosophical engagement” (HRDP 117). Such commentators evaluated Habermas’s review and Rawls’s reply as if it was the full on dispute that was billed, but never transpired. As Habermas himself later notes, explaining the disappointment, “I saw my role at the time rather as that of a reviewer who could expect a response from the author” (HR 283). It is unlikely that the mere prospect of a review of Rawls’s new book by Habermas would have provoked such high expectations and subsequent disappointment.
Is there, however, something to this widespread judgment? Yes and no. On the one hand, each had recently spent time and effort finishing lengthy and long overdue books setting out their mature political theory. They did not have much time to immerse themselves in each other’s work. Habermas read an early version of Political Liberalism, but was still thinking of it in terms of Theory. If anything, as he later acknowledged, the fact that the ideas of Faktizität und Geltung were still resonating in his head as he wrote his review may have stopped him gaining a proper understanding of the contours of Political Liberalism (HR 284). That said, Habermas is surely to be forgiven for that. For he only had the first version of Political Liberalism to go on. And, as Burton Dreben points out, Rawls’s ideas in the first version of Political Liberalism were not settled. For example he fails to distinguish consistently between a “political conception” of justice and a “comprehensive doctrine” until the second, paperback version, in 1996, after his exchange with Habermas (Dreben 2002, 320). On the other hand, the exchange was productive in that each forced the other to clarify their respective views. Habermas prompted Rawls to give a more detailed account of political justification, while Rawls forced Habermas to develop a more positive view of the role of religious discourse for social integration. To that extent, it seems both hasty and uncharitable to condemn the dispute as a failure.

In addition, according to Thomas McCarthy the eventual format of the exchange gave rise to several misconceptions. “[S]ince that time, the repeatedly discussed differences between Rawls and Habermas have been framed largely as differences within normative theory, more particularly, within what Rawlsians call ‘ideal theory.’” It is debatable to what extent Political Liberalism really is an exercise in ideal theory. Certainly, in design and conception it is less so than Theory. Nonetheless, McCarthy’s point is a good one: not noticing that the Habermas-Rawls exchange is in truth just a critical review of Political Liberalism and a reply, commentators inferred a similarity in object and approach between Habermas’s and Rawls’s political theories that did not begin to do justice to the differences in substance and approach of Between Facts and Norms and Political Liberalism. Perhaps they took this to be confirmed by Habermas’s observation that his critical comments in the 1995 exchange remained “within the bounds of a familial dispute” (RPUR 110/25) upon which so many critics have remarked. But the same point holds. It is clear when Habermas makes that remark in RPUR, and previously in correspondence with Rawls, that it refers specifically to their “exchange.” He is not saying that their respective works of political theory belong to the same family.

The 1995 Exchange

In Habermas’s opening salvo to the 1995 exchange, “Reconciliation through the Public Use of Reason,” after observing that he “shares” the intentions of Rawls’s project and “regards its essential results as correct” he offers three sets of criticisms:

I. on the design of the original position
II. on the idea of the overlapping consensus; and
III. on the relation between private and public autonomy and Rawls’s general approach.

I. Habermas argues that Rawls gets into trouble by constructing the original position as he does, building morally substantive considerations (of equality and impartiality) into the choice situation through the device of the veil of ignorance, which obliges the choosers to make impartial choices, while freeing them up from any moral pressure and allowing them free rein as merely rational self-interested parties, disinterested in others, who want the best outcome for themselves. He maintains this has three deleterious consequences.
A. The highly stylized parties in the original position, contrary to Rawls's explicit assumption, are unable to grasp the “highest order interests” of the actual citizens whom they model, and on whose behalf they choose (RPUR 111/27). Consequently, it is not the case that the principles chosen in the original position would be the ones that the actual citizens they stand in for would choose.

B. It misleadingly treats basic rights as primary goods, by imposing the prudential perspective of the economically rational chooser. It is thus unable to account for what he calls the “deontological meaning” of basic rights or principles of justice.

C. It does not guarantee impartiality of choice.

These claims are not new and have been much discussed in the literature, so I will not analyze them further here. Of interest is why Habermas kicks off what is supposed to be a critical review of Political Liberalism with these criticisms of the original position. It is puzzling since the original position, though it was the key argument of Theory, plays a much less prominent role in Political Liberalism. The idea in the later work is offered as a “conjecture” that the two principles of justice would be the most reasonable, if chosen from a list, by choosers in the original position from behind veil of ignorance. This conjecture is one of several possible “pro-tanto” or “freestanding” justifications of a political conception of justice, provided it can be shorn of its controversial philosophical assumptions and cleansed of all vestiges of Rawls’s comprehensive liberalism.

The question arises: Why begin a review of Political Liberalism thus? In my view, the best explanation is that this criticism is carried over from Habermas’s close critical discussion of the original position, which occurred in the late 1980s in the context of his dispute with the moral psychologist Lawrence Kohlberg over the construal of the moral point of view at Stage 6 of moral consciousness (MCCA; JS 220). We must forgo a detailed discussion of Stage 6 here. Suffice it to say that Kohlberg’s notion of “moral musical chairs” which provided his model for Stage 6 was directly inspired by Rawls's device of the original position. Of chief interest to us is that Kohlberg both generalizes and moralizes Rawls’s thought experiment, pressing it into the service of his own rather different agenda. As he writes,

Rawls’s theory is justificatory; it undertakes to show that certain principles of justice...are the ones that would be chosen...in the original position. In that case, Rawls claims, they are the right or true principles of justice. My psychological claim...is that something like his principles of justice are chosen by those at Stage 6, and they are chosen because they are reversible, or in better equilibrium than justice principles used at previous stages.

(Kohlberg 1981, 201)

Habermas made several telling criticisms of Kohlberg’s Stage 6, which eventually convinced the moral psychologist to change his view. For one thing, Habermas denied that Stages 5 and 6 are “hard” or “natural” stages, and contended that at this level the superiority of one stage over another consists solely in its degree of reflexivity and cognitive complexity and can therefore only be established by philosophical argument. For another, he contended that Stage 6 moral consciousness consists in the disposition to participate in a dialogical procedure with others, rather than to apply reason monologically, and that in this respect principle (U) offers a more adequate reconstruction of the moral point of view at Stage 6 (MCCA 170–175). It is these arguments that underpin Habermas’s claim in RPUR that “the meaning of the moral point of view cannot be operationalized in this way” (RPUR 112), i.e. in Rawls’s “monological” way (cf. McMahon 2002).
The upshot is that Habermas by this time has a heavy theoretical investment in these criticisms of Rawls. In addition, he tends to read *Political Liberalism* as a supplement, rather than as a revision to *Theory*. This is the best explanation for why he begins as he does. Nevertheless, by beginning with objections that had been formulated in the context of Discourse Ethics and aimed at the original position, Habermas derailed much of the subsequent literature, which gives undue prominence to the comparison between principle (U) and the original position, and to Habermas’s argument about the superiority of the former (see, for example, McMahon 2000, 2002). That comparison involves some philosophically interesting issues, for the question of how best to conceive of the moral point of view for the choice of impartial principles of justice is an important one in moral theory and in moral psychology. However, that question has little to do with Rawls’s *Political Liberalism* and throws little light on either the 1995 exchange or the broader dispute between them. In the introduction to *Political Liberalism*, Rawls criticizes his own approach in *Theory* and implicitly distances the new position from it: “In *Theory* a moral doctrine of justice general in scope is not distinguished from a strictly political conception of justice” (PL xvii). Rawls’s political conception of justice in *Political Liberalism* is thus not even “a conception of justice general in scope.” And it is certainly not a general principle of moral rightness, a procedure for establishing the validity of universal moral norms, which is broadly how it is construed by Kohlberg and how Habermas thinks of principle (U). No doubt, if the debate between Habermas and Rawls had been about how to “operationalize the moral point of view” in a general moral theory, a discussion of the relative merits of principle (U) and the original position as reinterpreted by Kohlberg would be central to it. But that is not what the 1995 exchange is about. Even if one considers the broader dispute between Habermas’s *Between Facts and Norms* and *Political Liberalism*, it is difficult to maintain that the salient points of comparison are principle (U) and the original position, since the former plays an ancillary role in *Between Facts and Norms*, although Habermas, as we will see below, is somewhat equivocal on that point. Nonetheless, the principle of democracy and the system of rights are the keystones to the theory of *Between Facts and Norms*.

II. Habermas’s second group of criticisms concerns the idea of the overlapping consensus. That said, at this point Habermas does not have a very clear grasp of that idea, and consequently tends to misinterpret it. Habermas objects that it is “unclear how the validity-claim of the theory is to be understood” (RPUR 119). And he suggests that either the idea of the overlapping consensus plays a “cognitive” role, like the notion of consensus and the associated idea of validity do in his own Discourse Ethics; or it plays a merely instrumental role in securing social stability, neither of which are strictly true. He then wants to press another dilemma. If the former is the case then this violates the method of avoidance; if the latter is the case Rawls fails to account for the epistemic and cognitive claims (or meaning) of democratic discourse.

There are actually several misconceptions here, which it is worth bringing to attention. To begin with Habermas claims that, as Rainer Forst puts it, Rawls “relegates the validity of the theory to the contingent fact of its being consonant with various comprehensive doctrines” (HRDP 157). I’m assuming that Forst here is not only reporting what Habermas’s view is, but also endorsing this view, for Habermas himself claims that Forst helped him to come to grips with Rawls’s theory: “I relied on his advice on first reading the unpublished manuscript of PL” (HRDP 295). Nevertheless, what the overlapping consensus of reasonable doctrines lacks, according to Forst and Habermas, and what it needs if it is to account adequately for the cognitive or epistemic meaning of democratic discourse, is “a common perspective on justice among citizens and a genuine moral consensus” (HRDP 157). In other words, it seems
that Forst and Habermas are thinking of the contents of the overlapping consensus of reasonable doctrines as something like the intersection set of a Venn diagram.

So the intersection set of reasonable Comprehensive Doctrines a. b. & c. might contain within it the political values <v, w, x, y & z>. Habermas suggests that they are all equally valuable, just because their value is due to their all occurring within the overlap of all the reasonable comprehensive doctrines, and because, consequently, all citizens will be minded to accept a political conception which is justified by some or all of them. To my mind, that is not how Rawls thinks of the overlapping consensus. For one thing, not every value in the intersection set will be a political one. There may be values or ideas that are shared by all citizens but that are not germane to the justification of a liberal constitutional democratic order (PL 139). So let’s assume that only <x, y & z> meet this second criterion of germaneness. For another, Rawls denies that the value of <x and y and z> only consists in their being shared by all reasonable doctrines (their falling within the overlap), and hence accepted by all reasonable citizens. By contrast, he contends that they are shared by all reasonable doctrines because they are “very great values and not easily overridden” (PL 169, 218). That is, they have intrinsic and not just instrumental value: indeed, they only have instrumental value because of their intrinsic value.

At the second stage of justification, Rawls claims, one asks whether Justice as Fairness is “sufficiently stable” (PL 141, 144). This means asking whether people have a “normally sufficient sense of justice” such that they comply with just institutions and laws, and also whether the political conception can be the “focus” of an overlapping consensus. On a common interpretation (to which I think both Habermas and Forst subscribe), consensus only comes into play the second stage, as a final test of the validity of the theory, after the freestanding justification has been worked up. The overlapping consensus, Habermas maintains, would in that case be merely an “index of the utility of the theory,” namely its “functional contribution… to the peaceful institutionalization of social cooperation” (RPUR 121–122) and it is no indication of what he calls the “validity” or the “epistemic meaning” of the political conception.

Now, as Quong points out, consensus is already in play at the first stage (Quong 2010, 3). After all, the political values and ideas, on which the freestanding justification of the political conception rests, are – as falling in the overlap of all reasonable doctrines – already held in common by all reasonable citizens. Qua freestanding, the justification avoids recourse to any comprehensive values that are not held in common and is thus indemnified from being reasonably rejected by any citizen on comprehensive grounds. The effect of the consensus operative at the first stage is that the freestanding justification is indemnified against reasonable rejection.
True the justification is pro-tanto, but that does not mean it cannot account for the epistemic meaning, or cognitive content of the political conception of justice; it means that it rests on a selection or subset of the full set of reasons on which the political conception can be supported. So, contra Habermas, consensus is operative at the first stage, and is apt to account for the epistemic meaning of the political conception.

Now let’s switch attention to the implied contrast, for Habermas contrasts what he takes to be the role of the overlapping consensus in Rawls to the role of consensus in his discourse theory. Habermas contends that the amenability to consensus of the political conception (or in his case – of a candidate norm) is internally related to its validity and thus a test of its rightness. Moreover, he takes rightness of norms to be analogous to the truth of propositions. That is because the notion of validity in discourse theory implies that if p is true, then it is amenable to rationally motivated consensus, and likewise if a norm is right, it is amenable to rationally motivated consensus, and that the validity claim to truth and rightness are analogous in this respect (Finlayson 2005). Note, however, that Habermas is implicitly drawing a comparison between Political Liberalism and Discourse Ethics, not Between Facts and Norms. The analog of truth, on his view, is the rightness (or justice) of moral norms. Thus, it turns out that it is the moral principle (U) and the norms it validates, that accounts for the “epistemic meaning” or cognitive meaning, which he claims that Rawls’s “freestanding” political conception of justice lacks.

It is true that Rawls does not make clear that consensus is operative at both stages of justification. And there seems to be an important ambiguity in Rawls’s use of the idea of consensus: an ambiguity between the notion of an overlap of reasonable doctrines, and a stock of political values held in common, on the one hand, and the agreement among reasonable citizens on a political conception of justice on the other. These are interrelated but ultimately different ideas.

Habermas’s third misconception is that an overlapping consensus for Rawls is test of the “neutrality” of the political conception, because what it means to call a justification “political not metaphysical” is that it is neutral with respect to conflicting worldviews. This does not capture the point acutely enough. Rawls denies that a political conception of justice is neutral in its effect. This cannot be because, as Rawls says, the political conception has the capacity “to shape those doctrines towards itself” (RH 145/58). It is only neutral in its aim, which is not to privilege any reasonable comprehensive doctrine, and in its grounds, since it may not appeal to any particular one, or group, of them.

III. In the final section of RPUR, Habermas levies three further criticisms. The first is that in spite of his claim of having balanced the liberties of the ancients and those of the moderns, in fact Rawls’s theory “generates a priority of liberal rights” and “demotes the democratic process to an inferior status”. As a consequence once the veil of ignorance is lifted, real citizens find themselves subject to “institutions and norms” that have been “anticipated in theory and have been institutionalized beyond their control.” Unlike in discourse theory, he argues, the citizens cannot understand the process of constitutionalization and the realization of basic rights as an “open,” “incomplete,” and “ongoing” process in the civic life of their society. (RUPR 128/42)

The second criticism is that Rawls conceives “political” as a technical term, with three different meanings. It describes a conception that is (a) neutral with respect to worldviews, (b) scope-restricted in its application to the basic structure of society, and (c) based on a small fund of values held in common. By contrast, discourse theory conceives of the political as those social domains that are regulated “by the legitimate means of positive and coercive law” (RUPR 130/43). This conception is both broader and more in tune with the actuality of political life in modern societies.
Finally, Habermas claims that his theory is modest in different ways (and by implication in a superior manner) to that of Rawls. His theory is modest in the sense that it can “leave more questions open because it entrusts more to the process of rational opinion and will formation” (RUPR 131/44), whereas Rawls's theory is modest in the sense that it deploys a “strategy of avoidance” which Habermas thinks is futile, and also a failure. Furthermore, he claims that where Rawls's theory is freighted by “substantive connotations” his own is developed in a “strictly procedural” manner (RUPR 116/32).

Rawls’s “Reply to Habermas” is over twice the length of Habermas’s initial review. Much of it takes the form of a patient elaboration and commentary on Habermas’s interpretation and criticisms of Political Liberalism, and a reiteration of his position on those points to which Habermas objects. Rawls gives very little quarter. He denies for example that his strategy of avoidance is unnecessary and that it fails (RH 150/63). He denies that his idea of stability (for the right reasons) is a merely functional or instrumental consideration that is not also part of public justification (RH 146/59). He denies that Political Liberalism leaves too much work to the political theorist and not enough work to citizens themselves (RH 153/65, 174/81). He denies that in Political Liberalism private and political autonomy are in unresolved competition, in contrast to discourse theory, which shows them to be co-original and internally related (RH 161–163/73–75). And he responds to the charge that his account is not “strictly procedural” by accepting that it is not and denying that it should be purely procedural, and stating that it combines substantive and procedural elements, just as does Habermas’s theory, albeit in a different way (RH 170/82).

However, there is one point on which Rawls offers not just a defense, but a significant development of his view, which is in his account of public justification. Here he outlines three different levels of justification and two different kinds of consensus, as supporting the idea of stability for the right reasons and legitimacy.

i. Pro-tanto justification is the justification for the political conception of justice offered on the basis of political values alone (RH 142/56).

ii. Full justification is carried out by individual citizens when they embed the political conception like a “module” in their respective comprehensive doctrines, such that their nonpolitical values and ideals support the conception worked up from the political values common to all reasonable doctrines (RH 143/56).

iii. Finally there is “public justification” whereby each citizen checks that every other reasonable citizen has successfully embedded the political conception of justice in their respective comprehensive doctrine. Rawls states that “the contents of these doctrines have no normative role in public justification” (RH 144/57). In other words, whether each individual can support the political conception from the perspective of their own comprehensive doctrine is a kind of social fact that each citizen is to ascertain for themselves. (So the third level of “public justification” does not require the kind of exchange of perspective, or ideal role taking, that, according to Habermas, moral discourse requires of its participants.)

Rawls also distinguishes two kinds of consensus. The first is a contingent convergence of values and interests, on which basis a skillful politician can forge a coalition to support a particular policy (RH 145/58). The second – which he calls a “reasonable overlapping consensus” – is essentially the consensus that consists in and results from the three-stage process of justification set out above, whereby, if successful, a political conception can be “supported...by reasonable citizens” even on the basis of their several comprehensive doctrines. (As I argued above, I believe there are two distinct ideas of consensus in play here: the overlap of reasonable doctrines on a shared fund of ideas and values, and the consensus of reasonable citizens on a political conception of justice.)
In addition to this elaboration of his conception of public justification, Rawls levels a significant criticism at Habermas’s discourse theory. “Of the two main differences between Habermas’s position and mine, the first is that his is comprehensive and mine is an account of the political and it is limited to that” (RH 132/47). And the first, Rawls argues, frames and stages the second difference, which is that between their respective devices of representation.

Rawls is right I believe to claim that the second difference, that between their respective devices of representation, is ancillary. After all, the dispute, insofar as there is one, is between their respective political theories, their conceptions of public justification, and their resultant ideas about democratic legitimacy and its relation to justice. As argued above, principle (U) does not stand at the center of Between Facts and Norms, and the argument from the original position is not central to Political Liberalism.

Rawls’s main objection is that Habermas’s theory is comprehensive and takes the form of a “logic in the broad Hegelian sense” (RH 139/53). The trouble with this objection is that Rawls’s conception of a comprehensive doctrine is polysemous, and includes religious, moral, and philosophical doctrines. Accordingly, there are at least three different ways in which Rawls’s chief objection can be construed. The objection can be that Habermas’s theory presupposes: (a) a worldview, religious or secular; (b) an actually existing morality or conception of the good; and (c) controversial philosophical theories of one kind or another, including moral theories such as utilitarianism or Kantianism. All of these three kinds of comprehensive doctrine have one feature in common, which is that they are all subject to reasonable disagreement, because there are mutually excluding alternatives that reasonable people might endorse. Perhaps that is why Rawls conflates them. He assumes that whatever specific sense of comprehensive doctrine is in play, the overall objection will be similar. Nevertheless, to assess the correctness of the objection, we will need to specify the sense in which the term is deployed.

The objection Rawls actually goes on to develop shows that what he has in mind is that discourse theory takes too many theoretical hostages to fortune – i.e. it is “comprehensive” in sense (c) – while Rawls’s theory “leaves philosophy as it is” (RH 49–51; cf. 48). That said, he could as well have developed the objection in either of the other two senses: that Habermas’s conception of legitimacy rests on substantive moral grounds or that it rests on a controversial secular worldview.

Evaluation of the Dispute

Habermas’s Criticisms

I. As regards Habermas’s first criticism, namely that Habermas’s device of representation is superior because dialogical, I believe it is misplaced as a criticism of Political Liberalism and is based on the comparison of Lawrence Kohlberg’s use of the original position, as a way of construing the moral point of view at Stage 6 of moral consciousness. In the debate about Stage 6 that comparison made sense, but it is not germane either to Political Liberalism and Habermas’s critique of it in the 1995 Exchange, or to the broader dispute between Habermas’s discourse theory of law and Rawls’s late political theory. Whether there is a meaningful difference between the two devices construed as alternative ways of choosing impartial principles in abstraction from differences in context, whether indeed the dialogical-monological distinction is even tenable, is questionable. For example, McMahon has argued in one article that the dialogical-monological distinction as Habermas conceives it is untenable, and in another that each theorist can
make use of the other’s device of representation (McMahon 2000, 2002). If one only considers their respective political theories, it is not clear whether Habermas’s discourse theory allows a greater role for public discussion in the generation of public reasons, nor, if it does, what advantage this is supposed to yield. Anyway the salient points of comparison with Rawls’s theory is not principle (U), but the principle of democracy and the resultant differences in their respective conceptions of legitimacy. And as we will see below, it is very unclear what the precise relation is between the principle of democracy as the reconstructed legitimacy-conferring procedure of democratic discourse, and the moral principle (U) as the reconstructed rightness-conferring procedure of moral discourse. 

II. This problem bleeds into the second set of objections Habermas makes, and muddies the waters there also. Recall that he tries to pin Rawls on either side of a dilemma. Either Rawls offers a merely instrumental justification for legitimacy in which case he fails to account for the “cognitive meaning” of democratically legitimate political norms or he accounts for their cognitive meaning but violates his own strategy of avoidance. Rawls counters that neither horn of the apparent dilemma is correct; since the justification of the political conception of justice is not merely instrumental, it has an essential moral dimension. And it is not external either, since the morality in question is a political morality and a core domain of the political.

The problem here lies not just with Habermas’s grasp of the Rawlsian position, but with his own position which forms the salient point of contrast. According to Habermas, moral agreement according to (U) must satisfy what I call “the validity requirement,” namely that every participant in a discourse agrees on a norm for the same reasons: “anything valid should also be capable of a public justification. Valid statements deserve the acceptance of everyone for the same reasons” (IO 86; Baynes, 2016; Finlayson, 2017). That is a very strong condition. It is not clear whether any moral norm can meet it in an actual discourse. Even as the counterfactual aim of an idealized discourse, it is very demanding. Habermas realizes this and acknowledges that under modern conditions the moral domain shrinks to a very small number of highly general norms (JA 91). Anyway, the point is that on his account the validity requirement of moral discourse forms the basis of the analogy between rightness and truth which Habermas argues is the basis of the epistemic meaning (or cognitive content) or moral discourse. Now whether democratic discourse must meet the same stringent condition, whether the validity requirement is also a necessary condition of any legitimate law is, on Habermas’s account, very unclear. The principle of democracy states that “only those statutes may claim legitimacy that can meet with the assent of all citizens in a discursive process of legislation that in turn has been legally constituted” (BFN 110). It contains several conditions but importantly does not state that an agreement on a law must meet the validity requirement.

In addition to these conditions, Habermas adds two more. In Between Facts and Norms he writes that legitimate laws “must also harmonize with... ethical principles” (BFN 99). On his view, ethical principles are not universally shared among citizens – Habermas is a pluralist about ethics and values – so what this condition means, it seems, is that the reasons that each citizen has to accept the laws do not contradict their various ethical value conceptions. There must be some congruence with the ethical values of the citizenry and the laws of the state. Habermas does not state how this condition is to be met in practice. (This looks prima facie rather like Rawls’s “full justification” whereby each citizen, for themselves, embeds the political conception of justice in their respective reasonable comprehensive doctrine.) The final condition, and the most important one for our purposes, is that legitimate laws must “harmonize with the universal principles of justice and solidarity” and that “a legal order can be legitimate only if it does not contradict basic moral principles” (BFN 99, 155).
I call this the moral permissibility constraint (MPC). Unfortunately in Habermas’s theory the important detail is lacking. MPC could mean that a law, if it is to be legitimate, may not be inconsistent with any extant valid moral norm. Some of Habermas’s statements lean in this direction (e.g. BFN 99). But it could also mean something much stronger, namely that there must be a positive relation of fit between all extant valid moral norms and any legitimate law. Kenneth Baynes, for example, takes Habermas to have the stronger position. “Citizens must simultaneously both presuppose and strive to articulate a basic political consensus...focused on the idea of a core morality that all citizens can endorse as valid for the same (publicly available) reasons” (Baynes 2016, 179).

If the latter stronger interpretation is correct, it looks like the validity requirement – that citizens reach agreement for the same reasons – may be a requirement of legitimate law too, on Habermas’s account. The necessary condition that every law subject to a “discursive process of legitimation” must satisfy is that it is amenable to universal assent for the same reasons, i.e. on the valid moral norms with which, Habermas maintains, all legitimate laws must cohere. Let’s assume this reading of Habermas is correct. Prima facie this is good news for Habermas. It bears out his assumption that his theory accounts for the epistemic or cognitive meaning of democratic discourse. But this victory comes at a high cost, for it implies that rationally motivated consensus in democratic discourse is just as hard to reach as it is in moral discourse. Consequently, the validity of political norms (their “legitimacy”) will be as hard to secure as the validity of moral norms (their “rightness” or “justice”). But for that very reason legitimate law cannot lift the burden of social integration from morality, which is what according to Habermas’s modernization theory it is supposed to do, once the moral domain shrinks to a narrow core of highly general norms. If, when faced with that unwelcome conclusion, Habermas opts for the weaker reading, while legitimacy becomes easier to come by on his account, he can no longer fairly claim that the principle of democracy and his account of legitimacy account for the ‘epistemic meaning’ or ‘cognitive content’ of democratic discourse.

Setting aside this difficulty for a moment and assuming (with Baynes) the stronger interpretation of MPC, what is the difference between Habermas’s view and Rawls’s? Prima facie, they are not so different. Public justification contains a moral core – for Rawls the political values that all citizens hold in common, and for Habermas the valid moral norms, or core morality. Public reasons are, both philosophers contend, moral reasons, which are indeed the same for all citizens. In addition, citizens must find reasons for endorsing the political conception (in Rawls) and legitimate laws (in Habermas) from within their several (and possibly discrepant) reasonable comprehensive doctrines (in Rawls) and worldviews or ethical self-conceptions (in Habermas).

So what is the difference? Well, for Habermas what is to be justified, or agreed to, in public justification, is not “justice as a political conception,” but rather all coercive laws or political norms in general. Moreover, for Habermas MPC is a side constraint on legitimacy exerted by general morality, whereas for Rawls the pro-tanto or freestanding justification of the political conception is worked up from values and ideas that comprise a subdomain of general morality, political morality – the politically germane contents of the overlap of reasonable doctrines.

So there are significant differences. But in fact, they do not bear out Habermas’s argument that his discourse theory accounts for, and preserves, the epistemic or cognitive meaning of democratic discourse (and the laws, policies norms, etc., that such discourse “validates”) while Rawls’s theory does not. That claim rests on a very strong reading of MPC, which is at worst thoroughly misleading, and at best subject to difficulties further down the line, as well as on a mistaken reading of public justification in Rawls.
III. There is more substance to Habermas’s third set of criticisms, in particular the view that Rawls makes the term ‘political’ into a technical term, far more restricted in scope than the everyday adjective is. Habermas is not alone in leveling these objections. On the one hand, perfectionist liberals like Joseph Raz have accused Rawls of confusing political philosophy, which aims at making true statements about politics and the political realm, with democratic politics, which aims at producing “stability for the right reasons” in a well-ordered society or something like that. And some defenders of Rawls have bitten that bullet by retorting that, unlike Platonic philosopher kings, in the real world philosophy has no independent contribution to make to political theory, where, just as in politics proper, philosophy must deploy the canons of political justification (Raz 1990, 10; Laden 2010). (I’m with Raz and Habermas on that point. What makes a law legitimate is rather different from what makes political theory justified, and justification in politics is rather different to justification in political theory.) On the other hand, others like David Enoch think that the public justification theorists’ preoccupation with the agreement or consensus that is supposed to result from justification, rather than with the substance of the justification, puts too much distance between the first-order matters of actual politics and the second-order contributions of political theorists. This criticism applies to Habermas just as much than it does to Rawls – because on his version the theorist’s job is too confined to the rational reconstruction of the practice of democratic argumentation.

Both criticisms apply to Rawls. There is a sense in which *Political Liberalism* is both too remote from actual politics and too political. On the one hand, most of the book is given over to a meta-theoretical discussion of the kinds of reasons or justifications that reasonable citizens must offer one another, and that the state or relevant government officials are to offer their reasonable citizens, for a political conception of justice that is used to gauge and shape the legitimacy of the basic structure of society. The account remains completely internal to the constituency of the reasonable, that is, even if unreasonable citizens benefit from citizenship through the rights and liberties they are granted, their unreasonable views are not taken into account in the justification of the laws and policies that define the political domain, namely the basic structure of society (Quong 2004, 320). To the extent that the unreasonable do not freely agree to be bound by the laws, they have to be “managed” or forced to comply with the laws that the reasonable agree to be bound by, although only so far as constitutional essentials and matters of basic justice are concerned. This means that *Political Liberalism* is too remote from real politics in two further respects. First, reasonable citizens are exempted from having to engage their unreasonable fellow citizens in argument. Second, their commitment to adduce only certain kinds of reasons (namely public ones) with which to articulate their political views, and advocate for legislation, devolves entirely on to a moral duty of civility.

On the other hand, Rawls’s theory is too political in that Rawls binds himself, and the justification of *Political Liberalism* to the same process, such that he, in advancing his theory, must act as if he were offering public reasons for legitimate laws or policies policy. But that demand seems both too much to ask of a political theory, and the wrong kind of demand to place on it.

**Rawls’s Criticisms**

Turning now to Rawls, what are we to make of his chief criticism that Habermas’s theory is comprehensive, while his remains political? In some recent important works, commentators sympathetic to Habermas, such as Hedrick, Heath, and Baynes, have sought to defend Habermas
from this charge (Hedrick 2010; Heath 2010; Baynes 2016). I think there is no straightforward defense of Habermas. Rawls’s objection involves a complex of different claims because of the triple ambiguity of the term “comprehensive doctrine,” on the one hand, and because of the implied contrast with a theory that is “political” in Rawls’s controversial sense, on the other. Before evaluating, we must disentangle them.

The first claim is that Habermas’s theory is comprehensive – in sense (c) above – in that it takes many theoretical hostages to fortune which reasonable people can reject. Note though, that Rawls’s criticism is contrastive. Habermas’s theory is criticized for being comprehensive and for not being “political.” So we should appraise the objection with and without the contrast. Ignoring the contrast for the moment, it is clearly better not to embroil a theory in needless theoretical controversies, which Habermas certainly does, by presupposing Discourse Ethics and the pragmatic theory of meaning and so forth. That said, in his reply to Rawls’s reply Habermas has a good retort to this criticism, namely that although a theory should strive to be “post-metaphysical” and “neutral” with respect to worldviews, “it does not follow that political theory can itself move entirely within the domain of the political and steer clear of stubborn philosophical controversies” (MW 93/10 77). Any political theory will make some controversial theoretical assumptions that reasonable people can reject. Take, for example, the very idea that political theory should itself conform to the canons of political justification (rather than philosophical justification). Habermas denies this too. That is, he denies Rawls’s contrasting claim that Habermas’s theory is not “political” on the grounds that that this is not an appropriate requirement for political theory. Habermas is right on both counts, and so has strong grounds on which to defend himself from both parts of this criticism.

Let’s look at Rawls’s criticism that Habermas’s theory is a “comprehensive doctrine in sense (a) above, i.e. as a religion or worldview. Of course, Habermas denies that his theory is a religion or worldview, or that it presupposes one. As a matter of interest, Charles Larmore has argued that Habermas’s conception of post-metaphysical as the appropriate way of doing philosophy under conditions of modernity amounts to a secular worldview (Larmore 1995, 63). But to my way of thinking, in the end this is just another way of saying that, because of its reliance on modernization theory, there are cultural and historical, and meta-theoretical, assumptions to Habermas’s theory here that can be reasonably rejected. The argument here is not so much about whether Habermas’s theory can prescind from making controversial assumptions, which he is right to deny, but whether the assumptions he in fact makes are justified.

Finally, Rawls could, though in fact does not, develop his objection in the following way: namely that Habermas’s discourse theory of law and the notion of legitimacy it puts forward rests on substantive moral grounds. I believe this is correct as a comment on Habermas’s conception of democratic legitimacy, even though Habermas takes pains to insist that the principle of democracy is derived from principle (D) and the form of law, and not from principle (U), and to insist that his conception of the political is thus fully autonomous, unlike natural law theory, which derives the authority of law in part from the normativity of an antecedent moral order (BFN 103–105, 104–105; BNR 78; cf. Finlayson 2016). This version of the objection holds because of the role of MPC as a necessary condition of legitimacy. The relevant difference between Rawls and Habermas on this point is that Rawls’s political conception of justice has moral grounds, but these consist in the ‘political’ values which form part of a subdomain of the moral, whereas Habermas insists that the moral norms that flow into the political system are general moral norms. It is the moral as such, not a subdomain thereof, that imposes side constraints on legitimate laws (BFN 109). Whether the MPC is a point against Habermas’s theory, or for it, is itself debatable. Many commentators take the view that Habermas’s discourse theory of rights and his conception of democratic legitimacy need more justificatory support from substantive moral norms (Larmore 1995; Forst 2011;
But the fact is that Habermas’s rejection of natural law theory, his insistence that his own theory of legitimacy is by contrast “morally freestanding” (BNR 80), and his construal of the autonomy of the political, all conflict with the actual role that valid moral norms play in discourse theory as side constraints on political legitimacy. It appears that Habermas himself has yet to decide which way to go on this issue.

Conclusions

In hindsight, it is clear that the early widespread judgment that the Habermas–Rawls debate was a “failure” was itself precipitate and mistaken. It rested in part on a conflation of the early debate and the 1995 exchange, or on confusion about the eventual format of that exchange, and, in either case, on a failure to see where the salient differences and points of dispute between their respective political theories lay. And while it is true to say that Habermas took a while to appreciate the novelty of Rawls’s project of Political Liberalism and that he misunderstood several of its important features, he was among its earliest critics and showed an unerring eye for its weak points.

Habermas was certainly right to criticize Rawls for narrowing the domain of political discourse and placing seemingly arbitrary restrictions on what can count as bona fide political justification, by restricting the constituency of reason-givers to reasonable citizens and choking back the fund of eligible public reasons to the political ideas and values already held in common by reasonable citizens. While this still allowed Rawls an elegant (if over intricate) public reason solution to the problem of finding a basis for legitimacy in the political domain, under persisting conditions of reasonable disagreement, to its critics it begged too many questions in favor of liberal democracy.

Habermas, though he had a broader and more empirically sensitive conception of the political, nonetheless put a lot of historical, empirical, and normative faith in modernization theory, and in associated processes of socialization; for example, in their ability to socialize individuals into moral agents at Kohlberg’s Stage 6, who as citizens have achieved a correlative political outlook that includes a commitment to democratic procedures and willingness to abide by their results.

As for the critical potential of Habermas’s political theory, he insists that it is still there. He contends that the moral point of view forms an immanent perspective “from which modern societies are criticized by their own social movements” (MW 110). However, such a contention is moot. A detailed account of how this is possible is lacking, and the relation between morality and legitimacy in Habermas’s political theory is unclear. Habermas still holds, contra Rawls, that substantive criticism of society on the basis of a conception of the good or the right is not appropriate for social theory. Principle (U) is supposed to be merely a reconstruction of a procedure of moral discourse, and it remains very unclear how a principle that serves “the descriptive purposes of rational reconstruction” of moral consciousness can simultaneously provide the basis of social criticism of a political society, albeit in the hands of citizens and social movements rather than the theorist. The suspicion remains, then, that Habermas’s political theory, much like Rawls’s, is more of a rich explanatory theory, than a normative and critical one. Or more charitably put, even if Habermas’s political theory is still a critical one, Rawls’s gives a much clearer account of its own normative foundations.

References


Further Reading

Baynes, K., Habermas (The Routledge Philosophers). New York: Routledge. (Good recent introduction to Habermas’s work with a discussion of Habermas and Rawls in Chapters 6 and 7.)


Larmore, C. (2008 [1996]) The Morals of Modernity. Cambridge: Cambridge University Press. (Historically informed analysis of morality and modernity with significant discussions of political liberalism and Rawls and Habermas.)