Rawls’ Stage of Full Justification and the Kantian Ideal of Autonomy

André Munro

From the onset of *Political Liberalism*, Rawls makes explicit that stability is a fundamental motivation of the new developments of his thought. Contrary to *A Theory of Justice*, the idea of the overlapping consensus outlined in *Political Liberalism* grows out of a deep concern for the stability of political justice and therefore gives a greater role for its justification to reasonable comprehensive doctrines, which power is nevertheless contained by the *proviso*.1 In the light of Habermas' criticisms, we will try to show that the legitimacy accorded to comprehensive doctrines in full justification arises from an “ethical-existential self-determination” conception close to the Kantian ideal of autonomy. Although Rawls' conception is a departure from transcendent idealism and comprehensive Liberalism, it is nonetheless inspired by the same ideal of self-determination, encouraging citizens to become “the author[s] of the law.” After exposing Rawls' view of the overlapping consensus justification and exposing the Kantian ideal of autonomy behind the stage of full justification, we will address Habermas' critiques pertaining to this stage of justification. Contrary to Habermas' contention concerning the lack of legitimacy of metaphysical doctrines as sources of philosophical validity, it will be argued that full justification serves primarily as a stabilizing force.

**Rawls' process of justification**

As already noted, Rawls explains the necessity of distancing himself from the comprehensive scope of the conception of justice he argued for in *A Theory of Justice* on the basis of political

---

1 The proviso is an injunction to the introduction of comprehensive doctrines in public debates. The proviso contrasts the public political culture from the background culture by requiring that conceptions of justice be able to be justified on political grounds alone. The introduction of comprehensive doctrines to support the political conception is legitimate as long as the latter remains free-standing in its justification base. See John Rawls, “The Idea of Public Reasons Revisited”, in Freeman, Samuels, ed., Collected Papers (Cambridge: Harvard University Press, 1999), p.591.
stability. Set out as a critique of utilitarianism, the systematic account of justice outlined in Theory is now conceived as too broad to account for the fact of pluralism. Rawls explains the distinction in this passage: “In Theory a moral doctrine of justice general in scope is not distinguished from a strictly political conception of justice. Nothing is made of the contrasts between comprehensive philosophical and moral doctrine and conception limited to the political domain” (Rawls 1996, xvii). Drawing from the idea of toleration originating in the end of the Religious Wars, Rawls seeks to develop a political conception of justice—in contrast to a comprehensive doctrine—which articulates a consensus among reasonable comprehensive doctrines, even metaphysical ones (ibid., xviii). One can see that Rawls and Habermas differ on the lesson to learn from the Enlightenment, the former putting the emphasis on toleration between comprehensive doctrines and the latter on the erosion of metaphysical grounds for political justification (Habermas 1998a, p.11).

For Rawls the process of justification, the wider acceptance of political, constructed doctrines, is threefold. First, the political conception of justice is formulated through practical reason in a solely political manner. This stage, referred by Rawls as pro tanto, attempts to develop a political conception of justice that is reasonable and valid independently of comprehensive doctrines. Indeed, the first stage projects a “free standing” conception of justice that expresses the “shared and public political reason” of the polity members (Rawls, op. cit., p.9). Because of the “shared” aspect of this conception, it cannot be dependent on comprehensive and controversial religious, metaphysical or epistemological theories; hence, it stands on political foundations alone (ibid., p.144). The fact that the political conception of justice must be “free-standing” in accordance with the proviso is also coherent with Rawls' departure from comprehensive forms of Liberalism like that of Mill or Kant (ibid., pp.195-199).

The second stage, known as full justification, lies at the core of the disagreement between Rawls and Habermas. In this stage, individuals reflect, in solo or in association with others, on

---

2 Philosophical, religious, moral, and so forth.
how a constructivist political conception of justice can be embedded and justified by his/her comprehensive doctrine (ibid., p.386). In full justification, the reasonable comprehensive doctrines gain much more influence than in the pro tanto stage. Even though the comprehensive doctrines cannot change the political conception in itself, they have the power to rank-order its importance relative to non-political values (ibid.). Moreover, they may even veto the pro tanto conception if it cannot be so embedded (ibid.). This is conceived as an important step, not because the insights of the comprehensive doctrines can lead to a more just conception of justice, but for the sake of stability. Again, we must remember that the full justification does not nullify the condition of the proviso: “Recall that a political conception of justice is not dependent on any particular comprehensive doctrine, including even agnostic ones. But even though a political conception of justice is free-standing, that does not mean that it cannot be embedded in various ways - or inserted as a module - into the different doctrines citizens affirms” (ibid., p.387). At the time of full justification the “principles of justice are already provisionally selected”, this second stage interacts between the political conception of the first stage and reasonable comprehensive doctrines to ascertain that the political conception can gain their support and brings about an overlapping consensus among an enduring majority of them (ibid., pp.65, 391). Failure to do so would put the acceptance of the political conception in jeopardy.

Finally, the members regroup in the public justification stage to get wide reflective equilibrium and justify the political conception as a consensus respecting their differences. In this stage, all reasonable members of society take each other in consideration as having reasonable comprehensive doctrines supporting the political conception without agreeing on their content (ibid., pp. 387-388). From an individual approval of the political conception, public justification completes the overlapping consensus by adding an inter-subjective perspective to the enterprise.

**Rawls and the Kantian ideal of autonomy**

The main reason why Rawls gives weight to comprehensive doctrines in
full justification is a belief in moral self-authentication. Although his political philosophy differs greatly from Kant's comprehensive account of ethics and search for *a priori* truths, this American philosopher still upholds the spirit of Kant's ideal of autonomy.

For Kant, the moral principles adopted by the rational autonomous person are not dependent on external authority (friends, church, society, etc), but genuinely arise from the individual willing the categorical imperative (Bell 2002, p.735). Positing a positive conception of freedom, Kant presents moral actions as those that are self-imposed under the principle of autonomy. In his *Fundamental Principles of the Metaphysic of Ethics*, Kant defines the principle of autonomy as follows: “always so to choose that the same volition shall comprehend the maxims of our choice as a universal law.” (Kant 1949, p.71). As H. J. Paton explains, this means that the autonomous agent acts in accordance with maxims “which can at the same time be willed as universal laws” (Paton 1971, p.212). Moreover, it is important to note that for Kant the concepts of free will and autonomy do not lead to a relativist and subjective stance toward the law, but on the contrary mean a self-imposition of the law as legitimate (ibid.). In the end, moral actions can only be moral insofar as the agent acts as if he was the author of the law. Lastly, Kant locates the origin of practical and theoretical reason in their own activity and opposes utilitarianism that is by definition heteronomous.³

Rawls rejects the Kantian conception of autonomy as supporting a form of Liberalism too comprehensive in scope. Faced with the fact of pluralism, Rawls cannot adopt the central position given by Kant to the ideal of autonomy, a liberal value ranked above and superior to all other norms and given a “regulative role for all of life”, since it would fail to respect the fact of pluralism (Rawls, op. cit., p.99). Moreover, Rawls rejects Kant's position that the order of values is constituted by the activity of practical reason. He regards this position, which he calls constitutive autonomy, as rooted in Kant's transcendent idealism. As such, and like other metaphysical views, it cannot stand as the foundation of a purely

political conception: “what is essential is that justice as fairness uses as basic organizing principles certain fundamental ideas that are political. Transcendent idealism and other such metaphysical doctrines play no role in their organization and exposition” (ibid. p.100). In this instance, Rawls is consistent with his method of avoidance. If political liberalism were to espouse constitutive autonomy, it would necessarily cease to be primarily political and would hence fail to bring about a fair political agreement among reasonable comprehensive doctrines.

Nevertheless, the process of full justification he outlines connotes a belief in the importance of autonomy in moral justification. In the exposition of the stage of full justification, Rawls does not explain if the comprehensive doctrines are autonomously affirmed or heteronomously imposed by the community. However, in his description of freedom, Rawls affirms that the citizens are also validly self-authenticating with regards to claims “founded on duties and obligations based on their conception of the good and the moral doctrine they affirm in their own life[…]” (ibid., p.32). In this sense, the role played by comprehensive doctrines in full justification is not a heteronomous factor but part of the individual autonomous will to the good.

However, as Habermas notes, Rawls extends the meaning of ‘autonomy’ and hence subordinates the justification of moral norms to the internalization and embedding of those within the life-history of the individual: “a person is free when he accepts authorship for his own life.” (Habermas 1998b, p.100). With Rawls, the question of autonomy is reversed: it is no longer whether the individual wills the political norm as a universal rule, but rather whether the political norm can be amenable to be willed by the individuals with respect to their ethical existence (comprehensive doctrines). In effect, Rawls transfers the meaning of autonomy from the plane of metaphysical consciousness to that of socially embedded consciousness. Contrary to Kant's notion of autonomy, Rawls' conception is immanent and political, and does not place individual consciousness above community and history. Daniel A. Dombrowski captures correctly the difference:

---

4 Rawls also supports “doctrinal autonomy”, meaning the ordering of political values on the basis of practical reason and in accordance with proper conception of society and person, see Ibid, 99.
For Kant it is the autonomous action of human being that explains the categorical priority of the right over the good, but it is crucial for Rawls not to fall into Kant's trap by making human autonomy radically disembodied. Causally determined nature, Kant thought, is transcended by human autonomy precisely because it is elevated above nature itself. Rawls avoids this trap (at least in PL, where the language of a Kantian noumenal self is dropped) by identifying autonomy with a political autonomy that is embodied and historical, and is not a pure consciousness that floats behind every one of its traits. (Dombrowski 2001, p.51)

This conception of embedded autonomy or as Habermas terms it “ethical-existential self-determination” is coherent with political liberalism's definition of the citizen. Far from an atomized or purely utility-maximizing individual, Rawls conceives the individual identity and thus moral will as rooted in both political and non-political commitments (Rawls, op. cit., p.30). In fact, all individual identity is assumed to have a comprehensive doctrine and non-public aims. Embedding the political conception into the comprehensive doctrine as a module is necessary to achieve a wide reflective equilibrium (Scanlon 2003, p.160). Making abstraction of these commitments and beliefs would be counter-productive, as the citizens might not even entertain the possibility of thinking outside the references of these comprehensive doctrines and moral convictions (Rawls, op. cit., p.31). Thus full justification, recognizing the ethical roots of individual identity, seeks to reconcile the comprehensive doctrines and the political conception of justice. Again, it is Rawls' concern for stability that brings in this ideal of symbiosis. Rawls’ hope is that, in the end, the political conception might shape the background ethical convictions and individuals’ comprehensive doctrines.

Besides having a political (rather than metaphysical) and embedded conception of autonomy, Rawls diverges from Kant on the process of normative integration. The overlapping consensus, which finally brings together the different comprehensive doctrines under a political liberal conception in public justification, is far more cosmopolitan than Kant's idea of the Kingdom of Ends.5

---

5 The notions of a pluralist consensus and of the collective deliberation entails in public justification are alien to the ideal of the Kingdom of Ends, which resemble Rousseau’s general will. See Robert Paul Wolff, *The Autonomy of Reason, A Commentary on Kant’s*
Rawls and Habermas's postmetaphysical argument
Habermas has different issues with Rawls' justification process. On full justification specifically, he believes Rawls' reliance on metaphysical comprehensive doctrines impoverishes the author's project, both epistemologically and in terms of political stability. Because of Habermas' post-metaphysical stance, this philosopher objects to the contention that religious and metaphysical philosophical doctrines might provide validity to political norms. He claims that a more consistently procedural account could provide an independent source of authority and hence save Rawls' theory. In his perspective, Rawls' concept of full justification is introduced as the expense of the realization of a truly valid and rational basis for consensus.

While Rawls proposes a free standing yet existential embedded conception, Habermas contends that only an independent conception, arising solely from procedural and practical reason, can be cognitively sound and socially integrating: “a political justice that stands on its own moral feet no longer needs the support of the truth of religious or metaphysical comprehensive doctrines.” (Habermas, op. cit., p.98). In this perspective, our metaphysical doctrines and ethical traditions have a role to play in reflective self-clarification and ethical authenticity but are not to be regarded as sources of moral validity.

Habermas clearly shares with Rawls the aims of political stability and autonomy. Nonetheless, he locates the solution for pluralist integration in the co-operative process of argumentation peculiar to ethical discourse. In his view, the possibility of the universalization and rational validation of moral norms is already premised in our communicative actions as demonstrated respectively by the principles ‘U’ and ‘D’ (Habermas 1990, pp.65-66). For Habermas, unrestrained argumentative speech has the capacity to bring an objective consensus on norms while recognizing the intersubjectivity of the different lifeworlds (Vermes 1994, pp.8-9).

Epistemologically speaking, Habermas might be right that an independent source of moral validity is needed. However, the stage of full

---

justification is not primarily a cognitive or epistemic tool, rather its goal is to assure stability in the context of reasonable pluralism. Habermas is correct in stating that Rawls, as another moral cognitivist influenced by Kant, seeks to analyze the possibility of impartial judgments on moral-practical questions (Habermas 1990, p.66). Yet, Rawls attempts to fulfill this impartiality ideal in the political (pro tanto) and public justification stages. The full justification stage, on the other hand, is first and foremost implemented to provide what Rawls calls “stability for the right reason”. Rawls makes explicit in this passage that full justification’s aim is social unity and the stability of the political conception of justice: “this basis of social unity is the deepest because the fundamental ideas of the political conception are endorsed by the reasonable comprehensive doctrines and these doctrines represent what citizens regard as their deepest convictions—religious, philosophical, and moral” (Rawls, op. cit., pp.391-392). Embedding the political conception of justice in the comprehensive doctrines aims at providing stronger acceptance and support (ibid., p.65).

Writing in response to moral contextualism, political liberalism seeks the support of comprehensive presence in an overlapping consensus as a condition of stability: “this overlapping consensus appears far more stable than one founded on views that express skepticism and indifference to religious, philosophical, and moral values, or that regard the acceptance of the principles of justice simply as a prudent modus vivendi […]” (Rawls, 1999, p.413). Rawls recognizes pragmatically that the citizens might not act as the “authors of the law” if the political conception opposes their existential-ethical horizon.

From the perspective of political stability, it could be argued that Discourse Ethics or other non-metaphysical (in Habermas' sense of the term) grounds of justification would be more stable. Yet, the idea of the end of metaphysics as a legitimate source of normative social acceptance is itself a very contentious hypothesis. The core of the disagreement between Rawls and Habermas on the stage of full justification relates to different views of the Enlightenment. Habermas, in a sense, may be said to be closer to Kant’s idea of the Aufklärung, which he understood as the

---

6 Reasonableness entails, among other things, accepting the Burden of Judgment and renouncing to use political power to repress other reasonable comprehensive doctrines (toleration), see John Rawls, Op. Cit., pp. 58-60.
detachment of our selves from the burdens of tradition. For Habermas, metaphysical and religious doctrines are not legitimate foundations for a social consensus, as they are themselves in need of justification in our post-conventional times. Rawls' theory moves within the Kantian tradition, but it puts the emphasis on tolerance rather than the total primacy of reason. Charles Lamore takes a very similar stand on the question of Modernity:

Habermas has misidentified the feature of modern experience that is crucially relevant to the basis of political association. The decisive element is not the waning of metaphysical and religious worldviews (though that has occurred). It is instead the recognition that such worldviews, as well as the recurrent postmetaphysical efforts to do away with them, are an enduring object of reasonable disagreement. (Lamore 1999, p.615)

Even if Habermas would provide sociological empirical works supporting the hypothesis of a decline in metaphysical doctrine’s authority, it is quite probable that the loss of traditional thought's legitimacy is a slow sociological process and that practical moral reason alone is not a sufficient ground for political liberalism to be stable in a pluralist society. Even in a perfect communicative setting, deeply religious yet reasonable people might not support a political conception completely divorced from their beliefs.

Conclusion
As we have explained, the stage of full justification is introduced by Rawls for the sake of political stability. Appealing to an extended conception of the Kantian ideal of autonomy, it brings to the forefront reasonable comprehensive doctrines into which the political conception of justice is embedded as a module. We only sketched Habermas' criticisms and it must be conceded that a more in-depth exploration of those would necessitate a more substantial analysis of his conceptions of Modernity and Communication Theory. In the end, however, the debate on full justification revolves around different uses of the word 'justification' (validity versus stability) and Rawls' pragmatic perspective on the implications of the fact of pluralism.
Bibliography


Vermes, Chayah, Validity and Validation in Kant’s and Habermas’ Moral Theories (MA Thesis, Concordia University, 1994).