Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.
Practical Reasoning and Transnational Justice: John Rawls's argumentation on justice in dialogue with Onora O'Neill's Kantian cosmopolitanism

Jill McArdle

Submitted for the degree of Doctor of Philosophy, at the University of Dublin, 2015
DECLARATION

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and is entirely my own work. I agree that the Library may lend or copy the thesis upon request. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Signed

Student Number 10261700

Date 24/04/15
Summary

This thesis evaluates two opposing accounts of justice and international justice that both claim Kantian foundations: firstly, the international account offered by the political philosopher John Rawls, and, secondly, the cosmopolitan account argued for by the moral philosopher Onora O'Neill. The primary focus of the research is the adequacy of their contrasting strategies of justification for principles of justice. The primary conclusion is that Rawls's justificatory strategy lacks the impartial perspective it claims and is fundamentally insufficient and unsuited to the actual global context with which modern liberal pluralist democracies find themselves. This conclusion is based on an analysis of Rawls's methodological starting points, which it is argued lead him away from universal-moral justification toward a context-dependent and partisan form of justification.

The analysis is divided into three parts. The first part addresses the sequence of Rawls’s argumentation on justice from his first book, *A Theory of Justice* (1971), to his later work, *Political Liberalism* (1993). The analysis proceeds by identifying Rawls's early methodological starting points. Of the key features identified, the idea of the social contract, which Rawls claims is derived from both Rousseau and Kant, is of particular importance to understanding the direction of his methodology. Rawls interprets it as a hypothetical agreement between members of a particular bounded society, as such his method and justificatory strategies are shaped from the beginning by a consideration of what hypothetical (and later actual) persons *would* agree to. This feature is fundamental to the method of argumentation of this thesis. Drawing on O'Neill's critique of Rawls, this thesis distinguishes Rawls's focus on what persons would agree to, i.e. the probability of consent or agreement, from what persons *could* agree to, i.e. the possibility of consent or agreement. The latter is understood to relate to the conditions for the possibility of action, which are explained as principles for action that respect the individual's capacities for action or agency.

This early emphasis on the probability of agreement affects the direction of Rawls's theory by orientating it toward conditions that motivate the actual acceptance of endorsement of principles. This is distinguished from principles that protect the capacity to accept principles. This orientation toward actual acceptance encourages the deployment of conceptions and ideas, such as the conception of the person, of the “bounded society”, and of “reasonableness” designed to encourage actual agreement on principles of justice. The analysis also draws on the work of the discursive ethicist Rainer Forst to examine the possibility of a universal-moral level of justification in Rawls's account.

In Part 2 this critique is applied *The Law of Peoples* (1999). The argumentation proceeds first by analysing whether Rawls's justificatory strategies are “objective” in the sense specified in *Political Liberalism*. Rawls's approach to practical reason, in
particular his conception of reasonableness (which it is argued grounds claims of objectivity) is analysed in light of his methodological emphasis on hypothetical and actual consent or agreement. His discussion of legitimacy is also examined and critiqued in light of O'Neill's arguments regarding the non-fundamental role of (actual or hypothetical) consent in justification.

The problematic role identified for consent in Rawls's justification is identified as the cause of the deficiencies in Rawls's account of international justice. In particular problems relating to the idea of a “people”, which requires an assumption of a legitimately shared conception of justice, and the role of boundaries, are addressed. Along with O'Neill, immanent cosmopolitan critiques of Rawls's focus on the bounded society are deployed to assess how appropriate his reasoning on justice is to the actual international context.

Part 3 sets out O'Neill alternative Kantian account of justice. It focuses on the “practical” and “political” nature of her interpretation, as her universalist use of these terms can be compared with Rawls's restricted understanding. Her account is analysed as a practical argument for constructing principles for a plurality of agents that focuses on rejecting principles that cannot be shared. Respect for the agency of humans as moral persons is identified as the most critical feature in O'Neill's account of justice. The justification for principles of justice is grounded in the supreme principle of practical reason. In this way her alternative Kantian account of public reason shares its foundation with the justification of principles of justice. Public reason is distinguished from private reason as reasoning that can be shared by a plurality of agents. This thesis concludes that, due to the necessary, practical and modal nature of O'Neill's account (along with its minimum assumptions about human nature), it offers a non-partisan and universal-moral basis for reasoning about justice and grounding obligations about justice. It is also argues that such an account is required in the current global context.
Acknowledgements

I would like to express my special appreciation and thanks to my supervisor, Prof. Maureen Junker-Kenny. I would like to thank you for your vital support during my time as both an undergraduate and a postgraduate student. Your advice and encouragement on my research has been invaluable.

I would like to thank Dr. Cathriona Russell of the Dept. of Religions and Theology for your help and support over the past couple of years, in particular these past few weeks. Further thanks also to Dr. Russell and both Dr. Andrew Pierce and Prof. Dr. Hille Haker for conducting my viva and providing excellent advice and feedback.

A very special thanks to my mother, words cannot express how grateful I am for your immense support and for the sacrifices you have made on my behalf. Thank you also to my aunt and cousins for their support. I would also like to thank all of my friends who provided much needed help and encouragement in times of stress.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CJ</td>
<td>Forst, Rainer. <em>Contexts of Justice: Political Philosophy Beyond Liberalism and Communitarianism</em>. Translated by J. Farrell. Berkeley: University of California</td>
</tr>
</tbody>
</table>
Table of Contents

Introduction........................................................................................................................8

PART 1
The sequence of Rawls's account of justice from A Theory of Justice
to Political Liberalism.........................................................................................................12

1.1. Rawls's A Theory of Justice..........................................................................................14
   1.1.1. Starting points: the concept of a bounded society as first articulated in A Theory of
          Justice ........................................................................................................................................18
   1.1.2. The Rawlsian interpretation of Kantian equality and its relation to the principle of
          reciprocity ....................................................................................................................................30
   1.1.3. The interpretation of Kantian autonomy in Rawls's constructivist account of justice
          35
             1.1.3.1. Freedom as modelled in the original position ..........................................................36
             1.1.3.2. Autonomy: the "highest-order" desire to act on principles of justice ...................37
             1.1.3.3. Congruence: autonomous action as an expression of the good .........................44
   1.1.4. Justification: considered convictions and reflective equilibrium ............................46

1.2. Political not comprehensive liberalism ........................................................................56
   1.2.1 Revised starting points and amended justificatory strategies ......................................60
   1.2.2. Freestanding conceptions: freedom and equality ......................................................64
      1.2.2.1. Rational autonomy: freedom to pursue ends .........................................................67
      1.2.2.3. Full autonomy: equality as a constraint on the freedom to pursue ends ..........69
   1.2.3. Freestanding conceptions: a political conception of justice ........................................69
   1.2.4. Justification and the political conception ......................................................................72
      1.2.4.1. Considered convictions in Rawls's account of practical reason ............................72
      1.2.4.2. Reflective equilibrium and citizen's considered convictions ...............................77
      1.2.4.3. Overlapping consensus and citizen's reasonableness ...........................................79
   1.2.5. Public Reason: the public basis of justification in liberal democracies ....................83

1.3. Conclusion ....................................................................................................................89

Part 2
The Law of Peoples: agreement as the foundation of international
law and justice.......................................................................................................................93
2.1. The conception of reasonableness as an "idea of practical reason": deficiencies in Rawls's fundamental justificatory strategy ................................................................. 97

2.1.1. A consensus of the "reasonable": the justification of Rawls's conception of practical reason .................................................................................................................. 99
2.1.2. Clarifications made in debate with other conceptions: Rawls's discussion of legitimacy and of actual and hypothetical agreement as forms of fundamental justification .................................................................................................................. 105

2.2. A critique of The Law of Peoples in light of the problematic status of Rawls's account of justification ............................................................................................................ 120

2.2.1. "Nonliberal" conceptions of justice and the endorsement of equality at the international level .................................................................................................................. 123
2.2.2. The neutralisation of plurality in the public sphere of liberal and decent societies 126
2.2.3. The idea of a consultation procedure to legitimise the affirmation of a nonliberal account of justice, and its problems .................................................................................................................. 130
2.2.4. Between justice and respect: Rawls's paradoxical account of decency as the outcome of his context-dependent strategy of justification .................................................................................................................. 135
2.2.5. A shared perspective on international justice? Decent regimes as agents in the original position .................................................................................................................. 145

2.3. The justification of state boundaries: exclusion and obligation in the context of global and transnational interaction ......................................................................................... 152

2.3.1. Objectivity and boundaries: the status of consensus in relation to justifying exclusion .................................................................................................................. 154
2.3.2. Obligations to distant others: immanent critiques of Rawls's state-centred account of international justice .................................................................................................................. 158
2.3.3. The inadequacy of the duty of assistance: idealised obligations for idealised societies .................................................................................................................. 170

2.4. Onora O'Neill on the scope of moral concern and duties of justice .......... 175

2.4.1. A critique of the negative duty not to harm as an incomplete account of obligation .................................................................................................................. 178
2.4.2. The scope of moral concern: the effect of connectivity on the structure of action .. 182
2.4.3. Recipience and agency: the rhetoric of rights and the practicality of obligation-centred accounts of justice .................................................................................................................. 187
2.4.4. The requirements of justice: preventing coercion, addressing inequality and limiting power .................................................................................................................. 189
2.4.5. The priority of justice over beneficence: assistance and obligation .................................................................................................................. 195

2.5. Conclusion ................................................................................................................................. 200
Introduction

It is the theme of this thesis to compare two approaches to international justice that are both located within the Kantian tradition, that of the political philosopher John Rawls, and of the moral philosopher Onora O'Neill. Despite their shared origins, both approaches arrive at markedly different conclusions on questions of transnational justice, in particular with regard to the scope of duties of justice. While Rawls restricts duties of justice to the domestic sphere and posits only a “duty of assistance” at the international level, O'Neill posits a “more or less” cosmopolitan scope for duties of justice, though her account differs from those who automatically equate moral universalism with a cosmopolitan scope of duty. This thesis argues that Rawls’s account reduces international obligations to a question of charitable beneficence and as such is inadequate to the moral demands posed by global interdependence. In comparison, O’Neill presents a convincing account of the scope of justice combined with a demanding interpretation of the requirements of justice, based on a recognition of the fragility of human agency and its vulnerability to coercive, deceptive and violent acts and practices.

Both accounts incorporate interpretations of Kantian conceptions of reason, freedom, equality and autonomy. However, their interpretations of Kantian morality differ in fundamental and critical ways. There are two major distinctions which I identify as crucial to understanding the different accounts of justice put forward by Rawls and O'Neill. One relates to the social contract, or to the role of agreement in justification. While Rawls interprets the social contract along the Rousseauian lines of a hypothetical or actual agreement in a bounded framework, O’Neill argues that Kant never made agreement or consent fundamental to justification; rather it was a secondary consideration restricted by the criterion of universalizability, a principle of respect for the agency of all humans as moral
persons. The other major distinction relates to the question of whether reason itself can supply the moral point of view. Rawls rejects what he interprets as Kant's transcendental claim, and instead seeks to ground justification in the objective perspective of reasonable persons as citizens. O'Neill, on the other hand, interprets Kant's assertion of the capacity of reason to supply its own conditions in a practical and modal sense: it is not a claim of the transcendental authority of reason, but a practical consideration of the conditions that must guide the thinking and acting of a plurality of finite and vulnerable agents who seek to coordinate with one another.

Rawls's emphasis on the hypothetical and actual acceptance or endorsement of principles of justice forces him to focus on empirical-psychological (rather than modal) considerations for achieving agreement. I argue that this ultimately leads his account away from universal-moral justifiability toward a consensus-based strategy of justification that is an insufficient foundation for principles of justice and public reason. The deficiencies of his approach mean that Rawls is unable to justify rights and obligations beyond the consensus and to manage the demands of ethical pluralism. This has serious implications for his account of international justice as it suggests a false dichotomy between, on the one hand, respect for "nonliberal" or foreign "peoples" and, on the other hand, justice in the sense of arguing for the universal validity of principles of freedom and equality.

In contrast to Rawls, O'Neill interprets Kant as offering a practical justification for the universal human capacity for autonomous moral reasoning. By rejecting shared premises as a starting point for justification, and seeking only modal constraints on the possibility of action, her account is able to offer a universal and non-partisan justification for rights and obligations. By justifying a principle of respect for the agency of all moral persons as a necessary and enabling condition for all reason and action, her account of practical, moral reasoning calls into question the apparent dichotomy between universal validity and respect for difference, and opens up possibilities for reasoning across boundaries and on a wide range of moral and ethical questions.
The contemporary global context is one of increasing integration at many levels, complex interactions across local, national and global levels and porous and permeable boundaries. Ethical diversity is therefore too deep and challenging to be able to ground an account of reasoning in the assumption of shared identity. Both Rawls and O'Neill link practical reason to public reason: the justification of principles of justice is tied to the conditions for public discourse. However, while Rawls interprets both “practical” and “public” in a highly restrictive sense that limits reasoning to the domestic national sphere, O’Neill’s Kantian account suggests that practical, public reasoning must be, in principle, accessible to “the world at large”. I argue that the account of justification and public reason she proposes is better able to criticise unjust practices and institutions based on universally accessible moral justification, to accommodate “reasonable” ethical pluralism and enable intellectual engagement and dialogue between cultures and traditions in the search for agreement on shared moral and ethical standards.

Philosophical questions of justification and the scope and limits of justice are not just academic matters, without relevance in both public and political spheres. Intellectual resources and philosophical clarifications make an important contribution to public and political discourse. In development work, for example, understanding the strategies of justification that underpin different approaches to justice can shed light on areas of disagreement and offer critical perspectives on categories like “aid” and “assistance”. The different forms of philosophical justification that are proposed offer legitimations for political actors to undertake particular actions: limiting justice to restricted domains offers justification for restricting or reducing aid to distant others, while universal-moral justification linked with a wide scope of moral concern demands political action on urgent issues like global poverty and immigration. How different thinkers define and interpret fundamental categories like justice, practical reason, rights and obligations can also help or hinder our capacity to identify violations of justice, for example by deploying either idealised or realistic accounts of human capabilities. It can also therefore help us identify deficiencies in the institutions and practices that guide the current status quo on justice. Furthermore, philosophical perspectives that endorse relativist or restricted accounts of reason offer
legitimation to those who seek to mask injustice and dispel criticism by appealing to ethical relativism and denouncing critique as the imposition of “western” values.

This thesis proceeds in three parts, addressing first Rawls’s account of justice, followed by O’Neill’s alternative cosmopolitan account. Part 1 traces Rawls’s account of domestic justice from *A Theory of Justice* to *Political Liberalism*, paying particular attention to his strategies of justification and his understanding of key Kantian concepts. Part 2 applies the considerations of Part 1 to Rawls’s account of international justice in *The Law of Peoples*, and attempts to explain how his limited approach to transnational justice is rooted in his early methodological decisions that reject Kantian starting points. It also introduces O’Neill’s alternative account and discusses her cosmopolitan critique of bounded theories of justice. Finally, Part 3 attempts to set out the critical aspects of O’Neill’s distinctive Kantian approach to justice. O’Neill’s account begins not by seeking the appropriate method for achieving political consensus, but is directed toward humans as “practical reasoners” who share a capacity for morality. This anchors the debate at a much deeper and more universal level than political-philosophical arguments that begin from the perspective of citizens. O’Neill interprets Kantian morality as “political” not in the sense that it is directed at a particular political polity, but in that it is concerned with the abstract social context in which all humans as practical reasoners find themselves.
PART 1

The sequence of Rawls’s account of justice from *A Theory of Justice* to *Political Liberalism*

Part 1 addresses Rawls’s work prior to *The Law of Peoples*, focusing on the sequence of his argumentation on justice from *A Theory of Justice* to *Political Liberalism*. His theoretical and methodological foundations are set out first via an analysis of certain key aspects of *A Theory of Justice* (1.1.). This is neither an exhaustive nor a comprehensive review, it is a selective account of particular features that relate to Rawls’s methodology. It focuses to a large extent on the “fixed ideas” he introduces as premises for his constructivist account, and on his particular interpretation of certain Kantian conceptions. His “fixed ideas”, in particular the conceptions of persons and society, are examined for their content and impact on the structure and outcomes of the constructivist procedure. Similarly the effect on his account of justice of his attempt to offer empirical accounts of conceptions of reason, freedom, autonomy and equality is addressed. Their empirical nature is critiqued as an unpromising alternative to Kant’s non-empirical understanding of reason and freedom. Finally, the dual strategies of justification are addressed: the original position and reflective equilibrium. Drawing on points made in Rainer Forst’s critique of Rawls, I suggest that both strategies are called into question by Rawls’s methodological foundations and his understanding of practical reason, which prioritise conditions for seeking *shared* premises and *actual* acceptance of principles of justice. This foundational emphasis on what persons *would* or *actually* endorse leads to strategies of justification that not only reduce Kant’s ethical vision from a universal-moral to a legal-institutional level, but also do not offer sufficient justification for Rawls’s principles of justice.
Turning to *Political Liberalism*, Rawls's proposal regarding ethical pluralism is reviewed as the impetus for his shift toward an emphasis on the "purely political" nature of his conception of justice (1.2.). His revised conceptions of reason, autonomy, freedom and equality are examined in light of their role in his amended justificatory strategies. These new strategies no longer seek to argue for a congruence between the right and the good but seek a neutral set of principles that can be endorsed by all "reasonable" comprehensive doctrines. The relationship between the stages of justification is examined, in particular regarding the role played by Rawls's understanding of practical (now political) reason, and the conception of reasonableness, in grounding justification. The basis for "public" justification is now presented as the citizens' perspective, and the conception of "reasonableness" grounds the objectivity of the citizens' perspective. It is the objectivity of the conception of reasonableness as a feature of practical reason that is later critiqued in Part 2 and ultimately rejected.

Finally, I briefly discuss Rawls's account of public reason as set out first in *Political Liberalism* (1.3.). For Rawls, an ethically pluralist society could not hope to achieve stable political agreement on matters upon which distinct ethical traditions differed. However it was his contention that a plurality of reasonable doctrines could come to agreement on a minimal yet sufficient set of liberal political principles. Given that, he argued, public reasoning would lead to conflict where different traditions attempted to argue for their particular vision of society, a neutral basis for reasoning must be sought. He therefore suggested that the values and principles of the *shared* liberal political conception of justice could serve as the basis for public reason. What is noted is that Rawls equates possible acceptance of principles with *probable* acceptance, a move that encourages the restriction of reason in the public sphere. The substantive nature of public reason is critiqued, along with its non-deliberative character. In concluding I argue that Rawls's foundational methodological decisions result in an inevitable theoretical move away from attempts to offer universal-moral justification toward context-dependent criteria for reasonableness and objectivity. This ultimately compromises his account's ability to offer a non-partisan foundation for practical reasoning about justice both within and beyond the bounded society.
1.1. Rawls’s *A Theory of Justice*

The opening section focuses on Rawls’s methodological starting points: the hypothetical social contract in the constructivist procedure, his anthropology, and relevant features of the bounded society (1.1.1.). It will highlight Rawls’s emphasis on the basic structure as defined by a system of rights and duties. Through his focus on this basic structure, and on a hypothetical social contract, the empirical framework with which Rawls grounds his theory of justice is apparent. Already this is a departure from Kant: from the level of universal-moral reasoning to the level of a legal-institutional framework for a bounded society, or at the very least an attempt to juxtapose the two. This legal-institutional emphasis embeds the notion of reciprocity within the groundwork of his theory of justice. While I suggest that reciprocity is not an appropriate fundamental philosophical principle for a theory of justice, it must be noted that the introduction of reciprocity was a welcome move away from the Utilitarian school, which struggled to prioritise or justify the status of the individual as a being with inalienable rights.

This section also focuses on the “fixed ideas” which Rawls argues are the necessary materials for a constructivist procedure, specifically the conceptions of person and society he puts forward. These “fixed ideas” are judged by Rawls to be necessary because he rejects Kant’s suggestion that reason itself has a law-giving capacity, that itself can “constitute” the moral law. With regard to the conception of the person, what is questioned is to what extent the impetus for justice is predicated on self-interest. While the sense of justice encourages moral persons to abide by the principles of justice, at this early theory stage self-interest seems to impinge on all aspects of the construction of principles of justice: it is uncertain how much shared interest is underpinned by self-interest, and within the original position itself, the selection of principles is based on fear and rational calculus. What is also noted is the role played by an anthropology of maximisation in grounding the motivation to seek principles of justice, along with the questionable assumption of mutual disinterest.

---

The nature of the bounded society is also discussed. This Rousseauian concept is introduced by Rawls again in order to "fix" the materials with which a constructivist procedure may be built. At this early stage I draw attention to the assumptions Rawls makes with regard to its closed nature. The nature of the bounded society later influences the contextualisation of public reason in *Political Liberalism* and further the anti-cosmopolitanism of *The Law of Peoples*.

I also draw attention to the use of (both) abstract (on Rawls's understanding) and ideal concepts, along with concepts introduced in order to bracket other possible features, and, finally, features that are taken by Rawls to be descriptive of actual societies (which he later describes as non-ideal). What is of note is the complex moulding involved in the construction of principles of justice; a complexity which introduces numerous problems that could potentially be avoided by the use of more strictly abstract premises, as has been suggested by O'Neill. Also, the juxtaposition of non-ideal with ideal features is of issue later with regard to his argumentation on international justice.

What emerges as the foundational standard to guide the establishment of principles of justice is the principle of reciprocity. Given the important relationship between reciprocity and equality, I next discuss the two together (1.1.2.). Equality is grounded in the two moral powers, a capacity for a rational plan of life and for a sense of justice (i.e. the capacity to limit one's pursuit of one's plan). This implies for Rawls that equal consideration is owed to those who can "give justice" in return; reciprocity underpins his understanding of equality. I note several implications of this, including that asymmetrical concern therefore appears to be ruled out (a crucial feature of Kant's ethics) and that deploying the criterion of reciprocity at such a foundational theory level risks undermining the justification of certain legal institutions, given that these are presupposed by reciprocity. I also suggest that it is his empirical, preference driven account of motivation that forces Rawls to make reciprocity theoretically fundamental; his reliance on the hypothetical and later actual acceptance of principles necessitates introducing criteria that increase the likelihood of actual acceptance.
Reciprocity grounds just action in Rawls's view, and Rawls follows Kant in describing this as autonomous action (1.1.3.). However his understanding of autonomy is deeply distinct from Kant’s understanding, given his empirical framework or reluctance to rely on Kantian conceptions of freedom, reason and agency. In order to avoid the attribution of noumenal freedom to moral persons, or indeed the agents in the original position, Rawls interprets this form of freedom as a freedom from natural and social contingency and builds it into the framework of the original position. The representatives then are credited only with an empirically defined freedom of choice in the selection of principles. Due to Rawls’s empirical account of agency and motivation, agents in his understanding cannot act aside from some form of desire and therefore he cannot conceive of the agent as capable of reasoning independently of their desires, hence the freedom that underpins moral reasoning must be embedded in the framework.

Rawls defines freedom as a proceduralised freedom from certain contingencies that facilitates an empirical freedom of choice (in opposition to noumenal freedom) and in this way it can be integrated with his account of autonomy. What is crucial to understanding the difference between his account of autonomy and that of Kant, following the work of Onora O’Neill, is that while Kant spoke of the autonomy of reason, Rawls speaks of autonomy of the self, meaning the individual self. And given that his framework links desire to action, he must connect the principles of justice to a form of desire that is internal to the self. This he attempts to achieve by describing the original position as a device by which the principles that express the true nature of the self as free and equal are constructed, therefore to act on those principles is to act out of a desire to express oneself as a free and equal moral person.

He also argues that it is not only a core element of the right, but also a crucial element of a person’s good, to express their nature as a free and equal moral person. This relates back to his justificatory strategy by showing how moral autonomy could, or would, come to be endorsed by moral persons, by arguing for a unity of the self that connected the right to the good. It was this element of his justificatory strategy that he later came to see as problematically tied to particular
conceptions of the good in a way that was incommensurable with an account that sought a path *between* conceptions of the good.

This argument from unity was not however central to his strategy of justification, rather his two fundamental strategies were, respectively, the constructivist justification, i.e. the original position, and what could potentially be described as a “contextual” justification, i.e. considered convictions in reflective equilibrium (1.1.4.). In analysing whether reflective equilibrium is grounded in a contextual form of justification I draw on Rainer Forst’s defence and critique of Rawls’s account of justice, as articulated in *Contexts of Justice*. I use his critique to highlight how the constructivist and “contextualist” approaches to justification are both rooted in the same source: the principles and ideas of practical reason. While Forst suggests that the contextual objection is misplaced because ultimately Rawls grounds all reference to considered convictions or the ideals of citizenship in a form of practical reason that he defends as having universal-moral justification, I suggest that the objection remains, though it must be directed at Rawls’s substantive and loaded account of practical reason, in particular to his *assumptions* about the content of the “ideas of practical reason”. I argue that because the content of his account of practical reason itself stands in need of justification, it cannot serve as the criteria for justification, nor hence as the standard for reasonableness. I also suggest that this methodological move is necessitated by Rawls’s particular starting point, specifically the focus on a hypothetical and then actual social contract; the emphasis on real agreement forces Rawls to build motivational concerns into each level of his justificatory strategy.

This I highlight in order to emphasise the potential advantage of an interpretation of Kant’s ethics, such as that of O’Neill, which does not attempt to replace Kant’s accounts of freedom, reason and agency with empirical concepts. The reliance on empirical concepts forces Rawls to load his account with substantive content that risks a slide into contextualism. Such a problem could be avoided by a focusing on the capacity of practical reason itself to provide the criteria for the establishment of principles of justice, and such a move can be made without relying on metaphysical assumptions.
1.1.1. Starting points: the concept of a bounded society as first articulated in *A Theory of Justice*

Rawls takes as his primary focus the justice of the basic institutions of a society. As his opening passage declares, "[j]ustice is the first virtue of social institutions, as truth is of systems of thought". Rawls follows Kant in arguing that the principles of justice for the basic structure of a society are neither intuitive nor self-evident, rather they must be constructed. What he attempts to construct is a hypothetical social contract, which he takes to be a continuation of Kant's constructivist approach. In contrast to Kant's transcendental approach, however, Rawls's approach to construction operates with empirical understandings of key terms, such as freedom and autonomy. Therefore from the beginning his aim is to show why empirical agents, who must always act from some desire or preference, and who each already have their own specified ends, *would choose*, and act on, principles of justice.

Construction requires materials, and Rawls provides these in the form of "fixed ideas", or conceptions of person and society that he believes will allow us to identify the appropriate restraints on just action. The following quote from the beginning of *A Theory of Justice* is instructive:

I shall begin by considering the role of principles of justice. Let us assume, to fix ideas, that a society is a more or less **self-sufficient association of persons** who in their relations to one another recognize **certain rules** as binding and who for the most part act in accordance with them. Suppose further that these rules specify **a system of cooperation** designed to advance the good of those taking part in

---

3 Though his approach is not exclusively Kantian: he describes himself as continuing the social contract tradition not just of Kant but of John Locke and Jean-Jacques Rousseau also. Importantly, however, Rawls does not specify the crucial differences between these approaches, which impact on the scope of his theory of justice, and on the understanding of the moral agent. The scope is Rousseauian, a bounded society, and his understanding of the negative rights of individuals derives from Locke.
4 Rawls's strategy at this stage is to appeal to rational choice theory, constrained by the appropriate impartial perspective, to argue for the selection of principles of justice by rational agents. Though he later abandons rational choice theory, we will see that it influences the description of the moral agent in *A Theory of Justice* and influences his account of the moral person through a reduction of certain inherent moral capacities. I will later argue, using the work of O'Neill, that such an account can be avoided by an alternative Kantian constructivist approach.
it. Then, although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as an identity of interests. (TJ, 4. Emphases mine)

From this we may discern the fundamental characteristics of a society in Rawls's conception: it is 1) self-sufficient as an association of persons, 2) rule- or principle-based 3) a system of cooperation, or, more specifically, a "cooperative venture for mutual advantage", and 4) there also exists both a conflict and identity of interests.

Taking the second and third features together, in this conception a society structures itself according to certain principles or rules, and these principles then in some way reflect the nature of that society as a cooperative scheme. A system of cooperation implies that the members of such a society will work together towards some specified end, and in this case the end proposed by Rawls is that of mutual advantage (among all citizens who are "free" and "equal"). A notion of cooperation itself does not entail that the specific end in question be that of mutual advantage, it only entails that there be some specific end. Therefore to conceive of society as a "cooperative venture for mutual advantage" is to introduce a set of apparently abstract concepts: one of social interaction as cooperation and the other of mutual advantage as a fundamental aim of society. It will have to be examined whether this will have the effect of ruling out in advance any specification of ends that do not fit the character of mutuality, or reciprocity, i.e. whether the rules to be decided upon are limited to those that reflect a specific notion of reciprocity.

With regard to the fourth feature, Rawls's assertion that both conflict and identity are features of societies generally, he explains the origins of both features thus:

There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts. There is a conflict of interests since persons are not indifferent as to how the greater benefits produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share. (TJ, 5)

---

3 The question of abstraction is one of contention, as we will see when we address O'Neill's objection to the Rawls' deployment of what he sees as abstract concepts. For now we will proceed with Rawls's use and interpretation of abstraction.
“Identity of interests” is understood as members’ shared interest in engaging in social cooperation, based on the recognition that deliberate isolation is inadequate for the pursuit of one’s interests, and therefore of anyone’s interests. Shared interest is grounded then in the self-interest of individuals. Conflict is understood to originate from the assumption that each individual, in pursuit of their interest, will “prefer a larger to a lesser share” of the goods available in their society. This conception introduces another set of apparently abstract concepts, one of society as marked by a certain conflict of interest, the other suggesting that a preference for larger shares of social goods is a characteristic of persons (who share a society) that is fundamentally relevant to their selection of principles.

What is evident at this stage is that, to a certain extent, an anthropology of self-interest and maximisation underpins the selection of principles of justice. And not only self-interest, but also mutual-disinterest, guides the agents who are to select principles of justice (cf. TJ, 13).^6

Self-interest is not, of course, the whole story in Rawls’s account: the members of the bounded society are also described as having a sense of justice and as willing to abide by principles of justice given the assurance that others will do likewise.7 However this sense of justice is not decisive in the construction of principles of justice. Rather, because Rawls believes that natural and social contingencies work against the rational selection of just principles by moral agents, the capacity for impartiality that guides just action is embedded in the procedure of construction, rather than in the agents themselves.

When it comes to the process of selecting principles of justice, the representative agents in question, the parties, are exclusively self-interested and mutually disinterested (with consideration of others relocated into the structure of the original

---

^6 “One feature of justice as fairness is to think of the parties in the initial situation as rational and mutually disinterested. This does not mean that the parties are egoists, that is, individuals with only certain kinds of interests, say in wealth, and domination. But they are conceived as not taking an interest in one another’s interests. They are to presume that even their spiritual aims may be opposed, in the way that the aims of those with different interests may be opposed.” (TJ, 13)

^7 “Persons are reasonable ... when ... 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)
position). While self-interest is a necessary condition of rational choice theory, to which Rawls initially subscribes, he introduces mutual-disinterestedness as a way of removing the possibility that altruism, or “benevolence” could be influential or decisive in the selection of principles of justice. As we have seen, Rawls is attempting to show why empirically described agents who must always act out of some desire would choose certain principles of justice. It is this struggle to explain why such empirically defined agents would rationally choose certain principles of justice that leads Rawls to separate agents’ rational and reasonable capacities in his initial constructivist procedure.

This also relates to Rawls’s reluctance to describe moral agents in terms of an innate, or noumenal freedom, as Kant did. This reluctance leads him to deny the capacity of free moral agents to reason about principles of justice and to reduce freedom in practical reasoning to a ‘freedom of choice’, defined in empirical terms. On Rawls’s account, agents are not reliably motivated by a preference for altruistic action to a sufficient extent for such a motivation for action to be included within a framework of rational choice. He therefore “abstracts” from altruistic motivation, or more accurately, he inserts an unvindicated ideal of mutual disinterestedness. However in doing so he also detaches the moral capacities of agents from the agents themselves. In his terms this means the bracketing of the agent’s capacity for a sense of justice until after the selection of principles of justice. Rawls sees the sense of justice as more than altruistic choosing of principles: it is the capacity to recognise the impartial (and therefore objective) validity of principles of justice.

---

8 “And if it is asked why one should not postulate benevolence with the veil of ignorance, the answer is that there is no need for so strong a condition. Moreover, it would defeat the purpose of grounding the theory of justice on weak stipulations, as well as being incongruous with the circumstances of justice”. (TJ, 149)

9 Onora O’Neill, “The Method of a Theory of Justice”, in Eine Theorie der Gerechtigkeit, ed., Otfried Höffe (Berlin: AkademieVerlag, 1998), p. 36. O’Neill comments, “Rawls assumes not only blanket ignorance of one’s own characteristics and social position, but also that parties to the original position are ‘mutually disinterested’, in the sense that they do not take an interest in one another’s interests, and specifically that ‘a rational individual does not suffer from envy’... All these assumptions are evidently false of ordinary persons, who know a lot about themselves and their social position, whose desires interlock and cross-refer in complex ways, and who may suffer envy. Abstraction from these realities would be a matter of assuming only that persons may or may not know much about themselves and their world, may or may not have cross-referring desires, and may or may not be envious. Idealizations assume that a particular ‘ideal’ predicate holds, and it is idealization (often misleadingly spoken of abstraction) that leads arguments into trouble.”
Practical Reasoning and Transnational Justice

and hence to act on such principles. He removes it from the original position but brings it back in when agents have to act on the principles of justice.

Rawls recognises that altruism (a selfless concern for the wellbeing of others) is not the same as justice in that his theory does not attempt to argue for principles of justice out of a concern for the wellbeing of others. However, the removal of altruism from the original position manifests as the bracketing of the sense of justice in the original position, implying that within the theory the two are to some extent co-identifiable. In the place of altruism Rawls inserts rational self-interest as the motivating factor, and therefore in the place of the sense of justice he inserts rational self-interested choice. Rawls employs empirically defined concepts and consequently the “sense of justice” is understood as an actualised as opposed to an inherent capacity and can therefore be identified with particular motivating preferences such as altruism. However, in contrast, a preference for the wellbeing of others is not identifiable with the Kantian understanding of morality, understood as an inherent capacity to reason about the appropriate principles of justice for a plurality of finite and vulnerable agents, regardless of specified motivating preferences such as altruistic tendencies or self-interest.

This highlights the different levels on which the two accounts operate. A Theory of Justice deploys empirical concepts of reason, freedom and action that lead to a methodological concern for the types of preferences that motivate action (i.e. self-interest versus altruism), whereas Kant’s approach affirms an innate capacity to reason practically about action in relation to other agents, whose agency is relevant to considerations of justice.

Taking the first characteristic of societies, that of self-sufficiency, we may take this to mean that such a society requires little or no assistance, support or interaction with other societies for survival (and also presumably for prosperity), for their own development and for interaction with other cultures. Indeed, when discussing the interaction between states, Rawls asserts that for his purposes we are to conceive of society as “a closed system isolated from other societies” (TJ, 8). Rawls does not justify this particular conception; while it is clear that there exists almost
no society that meets this condition, he takes it for granted that principles established for this case will prove relevant for existing societies (TJ, 8). This implies two assumptions about the nature of societies: 1) that international economic interaction and trade, as a component of a society's economy, does not affect (in a relevant way) the functioning of principles of justice for the basic structure of a society, and 2) more broadly, that all interactions of societies, whether political, cultural or economic, are irrelevant to the initial internal (or domestic) structuring of principles of justice.  

Though we will return to the question of closed and bounded societies when we examine *The Law of Peoples*, it is appropriate now to turn to Rawls's comments from *The Law of Peoples* on the role of national boundaries in theories of justice, because these comments reflect and articulate the reasoning behind his deployment of the "closed society" in *A Theory of Justice*. In response to the objection that current state boundaries are inherently arbitrary and therefore irrelevant to justice, Rawls responds:

> It does not follow from the fact that boundaries are historically arbitrary that their role in the Law of Peoples cannot be justified. On the contrary, to fix on the arbitrariness is to fix on the wrong thing. In the absence of a world-state, there must be boundaries of some kind, which when viewed in isolation will seem arbitrary, and depend to some degree on historical circumstances.

According to Rawls, it is impractical to adopt a position on justice that rejects the necessity of taking some boundaries as given when reasoning about just principles and consequently about courses of just action that are to apply to the world as it is. Further, Rawls connects the justification and legitimacy for international

---

10 Note that these assumptions are not identical with the suggestion that international interaction is always irrelevant to the internal justice of a society; they note only that principles for the domestic society do not need to take into account *a priori* the types of economic relationships societies have with each other; international interaction is only irrelevant when constructing the principles of justice. It leaves open the possibility that international interaction can have significant effects on the continued just functioning of a society.

11 Although the discussions in *A Theory of Justice* and *The Law of Peoples* have differing focuses, on the abstract conception of the domestic society and on the conception of the international community as made of "peoples", the comments may be taken together, as Rawls's reference to a bounded society remained constant throughout.

boundaries to the role of government as caretaker for a territory, as "the effective agent of a people as they take responsibility for their territory and the size of their population, as well as for maintaining the land's environmental integrity" (LP, 8). Rawls claims that the allocation to a particular society of all benefits and burdens attached to a particular territory is essential for the maintenance of that territory:

Unless a definite agent is given responsibility for maintaining an asset and bears the responsibility and loss for not doing so, that asset tends to deteriorate. On my account the role of the institution of property is to prevent this deterioration from occurring. (LP, 8)

A certain agent must be given responsibility for the care of particular lands, resources and institutions, and the institution of property, and therefore of ownership of lands and resources, is necessary to incentivise the agent in question to maintain its territory. If the agent is not entitled in this way, if robust property rights relating to land and resources are absent, the incentive to care for the territory is lost, and deterioration is inevitable. A consideration worth noting at this point is that it seems that whether (legally institutionalised and coercively enforced) boundaries may be taken as theoretically or philosophically foundational when thinking about justice is a separate question to whether human agents, individual or collective, can be incentivised to manage assets whose benefits they do not have sole entitlement to. One final comment from The Law of Peoples I will note only briefly, as it will be taken up further in relation to accounts of distributive justice. Rawls writes:

I believe that the causes of the wealth of a people and the forms it takes lie in their political culture and in the religious, philosophical and moral traditions that support the basic structure of their political and social institutions, as well as the industriousness and cooperative talents of its members, all supported by their political virtues. (LP, 108)

This statement is particularly significant as it further reinforces the justification for Rawls's initial limiting in A Theory of Justice of the conception of society to that of an isolated society. It is slightly more evident now why we may conceive of society for the time being as a closed society; all the essential elements of justice, along with the wealth that is to be (justly) acquired and distributed, can be
secured without any need for continued external assistance.\textsuperscript{13} Rawls makes these comments directly as a response to arguments for global principles of distribution; his focus when he considers national boundaries here is the distribution of wealth across these boundaries.

One other consideration to note at this point, by taking the comments from \textit{A Theory of Justice} and \textit{The Law of Peoples} together, is that the conception of society in question is intended as something bearing similarity to a nation-state.\textsuperscript{14} So another fundamental feature of this conception of society, which is to serve as a "fixed idea" for Rawls's first book, is that the basic form of society in question is a state. One corollary of this position is that it therefore cannot be any other form of society, either subsidiary to or aggregate of nation-state, such as a municipal or global community.\textsuperscript{15}

Returning to \textit{A Theory of Justice}, once the conception of a society has been "fixed", all these characteristics taken together lead to a particular conclusion on the origin and nature of principles of justice: these principles are necessitated by the preceding premises. Once we conceive of society as a self-contained cooperative enterprise between "mutually disinterested" (TJ, 13) members who come into conflict over claims over the distribution of resources, while agreeing that cooperation is preferable to competition and antagonism over said resources, we arrive at the conclusion that the appropriate principles of justice "are the principles of social justice: they provide a way of assigning rights and duties in the basic institutions of society and they define the appropriate division of the benefits and burdens of social cooperation"(TJ, 4). On this understanding social justice is comprised of a system of rights and duties, with a division of benefits and burdens as a form of distributive justice.

\textsuperscript{13} States will not require continued external assistance, although some states may require \textit{temporary} assistance in reaching an adequate level of development. Also, this is not equivalent to the claim that societies are immune from \textit{negative} external influence, though what constitutes "negative" and what constitutes "external" are perhaps a separate question and matter for further consideration.

\textsuperscript{14} For our purposes though I take peoples to be relatively equivalent to states, although Rawls argues for some crucial differences. See Rawls, \textit{The Law of Peoples}, pp. 23-29.

Further comments help elucidate the connection between this particular account of domestic justice and the conception of society, and provide us with a more adequate understanding of the nature of this conception of justice. Rawls writes:

\[W\]hile men may put forth excessive demands on one another, they nevertheless acknowledge a common point of view from which their claims may be adjudicated... If men's inclinations to self-interest make their vigilance against one another necessary, their public sense of justice makes their secure association together possible... the general desire for justice limits the pursuit of other ends. (TJ, 5)

The role of the principles of justice is then to mediate the claims of individuals, and the system of rights and duties proposed by Rawls is one that begins (and must begin) with rights, from which duties are to be inferred. The necessity for a system of justice based fundamentally on the claims of individuals stems from a specific conception of interpersonal conflict as based on "excessive demand" and "inclination to self-interest". There are several aspects of this conception that I would draw attention to at this point. The first is that justice is defined as a limit feature on the self-interest of rational actors: justice exists not as a goal in itself but rather as a legal-regulative circumscription of self-interest, which all recognise, through the sense of justice, as necessary and right. The second is that the justification offered by Rawls for a rights-based system of justice originates in the understanding of conflict found in the conception of society. Although Rawls assumes an anthropological "sense of justice", it is this assumption of conflict over primary goods that leads Rawls to ground social justice in the rights, or claims, of individuals. Thirdly, therefore, duties become secondary and contingent; the inference of duties from rights has this effect. The fourth is that such a position rules out in advance the possibility of starting from any alternative agent perspectives, ones that do not take the rights or claims of individuals as exclusively fundamental.

Rawls introduces one final characteristic of his conception of a society, that of well-orderedness. This implies that a society is "not only designed to advance the good of its members but... also effectively regulated by a public conception of justice" (TJ, 4). With the introduction of the concept of "well-orderedness", of public
adherence to principles of basic social justice, Rawls moves away from (what it is claimed is) an abstractly descriptive account of societies to a now normative account of what constitutes a just society. With this the theory of justice presented in *A Theory of Justice* locates itself within “ideal theory”, where “[e]veryone is presumed to act justly and to do his part in upholding just institutions” (*TJ*, 8). Rawls acknowledges that this moves the method away from a more descriptive account, stating that he is concerned with “strict” and not “partial compliance theory”, with the assumption of strict compliance functioning as a normative ideal to be aimed at (*TJ*, 8).

The justification he offers for such a move is the assertion that issues relating to “nonideal theory”, such as civil disobedience or institutional injustice, can only be addressed once prior questions relating to the “nature and aims of a just society” have been settled (*TJ*, 9). Based on this assertion Rawls leaves aside almost entirely questions related to nonideal theory, or to “partial compliance”. It is only in later *The Law of Peoples* that he returns to nonideal theory, with the consideration of the status of “peoples” that are not “well-ordered”.

Rawls initially provides us with what he takes to be the general qualities of societies relevant for justice: that they are cooperative ventures for mutual advantage, marked by both conflict and identity of interests. In addition to these, he then proposes two features that are not intended to be understood as general empirical features of most actual societies: that they are closed and well-ordered. The addition of “well-orderedness” can be explained by the move from apparently descriptive claims to aspirational normative claims about just societies. It is not so readily apparent however why he introduces the notion of a closed society. It is not descriptive, given that no actual society meets this condition. It is not a normative claim either; it is not intended to suggest that self-sufficient societies achieve greater justice than internationally engaged societies. Instead, it is a claim introduced in order to *bracket* certain actual features of societies that Rawls considers to be irrelevant to domestic justice. Given that Rawls believes domestic societies succeed or fail almost entirely through their internal political institutions,
the question of transnational justice is therefore not a factor that should influence
the initial construction of the principles that guide domestic political institutions.

There is an interesting comparison between the use of supposedly abstract
concepts in this case and in the case of the original position in A Theory of Justice.
In the case of international interaction, we have a factor that is not believed, in
actual fact, to influence to a relevant degree the justice of domestic social
institutions. In the original position on the other hand, the bracketed information
is thought, in actual fact, to influence the justice of domestic social institutions.
Rawls's concerns for the original position are evident when he comments that
"institutions are just when no arbitrary distinctions are made between persons in
the assigning of basic rights and duties" (TJ, 5), the assumption being that because
arbitrary distinctions can influence the design of systems of rights and liberties,
they must be bracketed.

Both types of features are in some sense irrelevant for Rawls; either because they
have no real influence, or because the influence they have is deemed arbitrary.
Rawls evidently considers both forms of bracketing to be suitable interpretations
of the method of abstraction: bracketing both information that does not, and
information that should not, influence the construction of principles for a domestic
society. We will explore this further when we look at O'Neill's objection to the use
of abstraction in Rawls' theory, beginning in his first book but also continuing into
his writing on international justice.

In summary, the conception of society presented in A Theory of Justice comprises
several types of features. Firstly, there are the characteristics of societies that
Rawls takes to be true, abstract and relevant to domestic justice; that they are
systems of cooperation, that the aim of cooperation is mutual benefit, and that
members get into conflict over the distribution of benefits and burdens while
rejecting antagonism as an ongoing condition. Then there are artificial
characteristics; those which are not true in any sense, but whose inclusion
represents the bracketing of certain other features that are true, abstract, yet
irrelevant to justice. So the notion of the closed system can be seen as a device
deployed to bracket the assumed irrelevance of inter-societal interaction. Finally there are the normative claims, specifically the notion of society as well-ordered. This claim locates this account of justice within ideal theory. The conception of society presented to us in *A Theory of Justice* therefore comprises a triad of, firstly, (apparently) abstract features, secondly, artificial features and, finally, ideal features. While the use of abstraction is generally uncontroversial, I would draw attention to three accounts on which objection may be raised to Rawls’s particular approach. The first is whether Rawls’s method of abstraction is in fact an unproblematic type of abstraction. As we will see, O’Neill suggests it is not. The second is whether the bracketing of certain true but supposedly irrelevant information is based on false assumptions and therefore detrimental to the theory; is autarky a suitable starting point for a domestic theory of justice, and what does it imply for international justice? The third relates to ideal theory, and to whether the amalgamation of ideal features with abstract or nonideal features is a promising approach to constructing principles of justice that are to apply to the world as it is.

These features together lead to the initial impetus for the construction of principles of justice; such principles allow the members of a domestic society to be able to cooperate together for mutual advantage. Given that the sense of justice is not *decisive* in the selection of principles of justice, and that inclination to self-interest will create antagonisms, these principles will create an impartial space where the claims of all members may be adjudicated, where “the rules will determine the proper balance between competing claims” (Tj, 5). A crucial requirement of this impartial adjudication is that it be *fair*. While condition is already manifest in the notion of mutual advantage, and in Rawls’s concern to bracket information and features irrelevant to justice, we will see more clearly the origins of fairness or impartiality as the foundational standard of justice in Rawls’s theory when we look at how he interprets and attempts to integrate Kant’s account of justice – which is based on moral grounds, rather than self-interest, and is of cosmopolitan scope – into the method of *A Theory of Justice*. 
1.1.2. The Rawlsian interpretation of Kantian equality and its relation to the principle of reciprocity

Kantian equality, according to Rawls, is a form of equality owed to moral persons. Such persons are understood to have two basic capacities:

[F]irst they are capable of having (and are assumed to have) a conception of the good (as expressed by a rational plan of life); and second they are capable of having (and are assumed to acquire) a sense of justice, a normally effective desire to apply and to act upon the principles of justice, at least to a certain minimum degree. (TJ, 505)

All persons with the requisite capacities are to be considered equal and are owed equal justice. This does not imply that all persons must have these capacities in equal measure; rather persons will exhibit these capacities to varying degrees. What is of importance for Rawls is firstly that this capacity be understood primarily as a “potentiality”, and secondly that it be understood as a minimum requirement. Regarding the former, this “potentiality” is one he considers to be “ordinarily realized in due course”; we may take for granted that the majority of persons realise their capacities at some point in their lives. He describes the lack of “potentiality” as a “defect or deprivation”, considering these cases to be aberrations from the norm. Later, in Political Liberalism, Rawls further elucidates this by adding that equality is conditional on the capacity to be “normal and fully cooperating members of society”.\(^\text{16}\)

Regarding the latter point, Rawls stresses that these capacities need not be of a highly developed nature, nor is it appropriate to make interpersonal comparisons that evaluate the degree to which the capacities are developed; that they are capable of being realised above a basic threshold is all that equality requires:

\[^{16}\text{Rawls comments: "Here we appeal to the fundamental idea of equality as found in the public political culture of a democratic society just as we did with the three ways in which citizens regard themselves as free persons... We noted this idea in saying that citizens are equal in virtue of possessing, to the requisite minimum degree, the two moral powers and the other capacities that enable us to be normal and fully cooperating members of society. All who meet this condition have the same basic rights, liberties, and opportunities, and the same protections of the principles of justice." John Rawls, Political Liberalism (New York: Columbia University Press, 1993), p. 79.}\]
First of all, the simplicity of the contract view of the basis of equality is worth emphasizing. The minimum capacity for the sense of justice ensures everyone has equal rights. The claims of all are adjudicated by the principles of justice... Nor does equality presuppose an assessment of the intrinsic worth of persons, or a comparative evaluation of their conceptions of the good. *Those who can give justice are owed justice*. (TJ, 510. Emphasis mine)

Furthermore, he adds that “[b]y giving justice to those who can give justice in return, the principle of reciprocity is fulfilled at the highest level” (TJ, 511). These comments further elucidate the notion of a minimum threshold; what is relevant for such a threshold is the capacity to partake or cooperate in a just social scheme. Given that the scheme is one of cooperation for *mutual* benefit, the relevant principle is that of reciprocity; we can each *contribute* to the cooperative scheme to a sufficient degree, and it is on this basis that we are owed justice. For Rawls, once we establish equality in this sense, the principle of reciprocity follows directly. Equality here seems to be articulated as an empirical concept, in that it interprets the shared sense of justice as *actualised*, rather than as inherent in human beings, as it is in Kant. Rather than reflecting an inherent capacity to act from a sense of justice, it appears to be reduced to an actual capacity to participate in cooperative schemes, motivated by the assurance that others will do likewise.

Rawls's reductive interpretation of the Kantian conception of equality underpins the notion of reciprocity in *A Theory of Justice*. All relationships and interactions between members of a well-ordered society then are regulated by the principle of reciprocity. Indeed early on in his first book Rawls indicates that equality should

---

17 In comparison, O'Neill has argued that reciprocity is not a fundamental criterion for distinguishing the scope of moral concern, nor is the range of consideration dependent on defining personal attributes in the form of particular capacities. Rather than attempting to define the moral subject, she suggests that it is the other as *agent* that is relevant. According to O'Neill, "when agents commit themselves to the assumption that there are certain others, who are agents or subjects with these or those capacities, capabilities and vulnerabilities, they cannot coherently deny these assumptions in working out the scope of ethical consideration to which they are committed. Commitments to others' ethical standing are taken on as soon as activity is planned or begun: what is needed is a procedure for working out what these commitments are in a given context". Onora O'Neill, *Towards Justice and Virtue: A Constructive Account of Practical Reason* (Cambridge: Cambridge University Press, 1996), p. 100. I will return to these considerations in Part 2 when considering the scope of moral concern in *The Law of Peoples*, as it seems that O'Neill's account offers a more promising and practical approach to considering which agents are owed justification within or beyond national borders.
be so understood when he equates “people who view themselves as equals” with those “entitled to press their claims upon one another” (TJ, 14).

It is helpful to briefly review how Rawls understands the notion of reciprocity in *A Theory of Justice*. Firstly, it is defined as “social cooperation among equals for mutual advantage” (TJ, 14). This we saw when looking at Rawls’s conception of society. Cooperative advantage in *A Theory of Justice* is equivalent to “reciprocal” advantage (TJ, 178).

Of primary concern here is that (in contrast to Utilitarianism) no one must believe they are bearing a disproportionate share of the burden, and this includes both the least and most advantaged.\(^{18}\) This for Rawls was the mark of a stable society. Stability is brought about via a shared understanding of society as grounded in a principle of reciprocity; each member has an expectation of receiving in return something comparable to what they have contributed (in terms of compliance with the principles of justice and claims on primary goods). *We willingly act on just principles because we expect others will do likewise.*\(^{19}\) According to Rawls, persons with divergent conceptions of the good could, or (more precisely) would, come to endorse justice as fairness if they had a legitimate expectation that the scheme was

---

\(^{18}\) Rawls writes of the contrast between Utilitarianism and justice as fairness: “In fact, when society is conceived as a system of cooperation designed to advance the good of its members, it seems quite incredible that some citizens should be expected, on the basis of political principles, to accept lower prospects of life for the sake of others... Looking at the question from the standpoint of the original position, the parties recognize that it would be highly unwise if not irrational to choose principles which may have consequences so extreme that they would not accept them in practice. They would reject the principle of utility and *adopt the more realistic idea of designing the social order on a principle of reciprocal advantage*”. (TJ, 104. Emphasis mine)

\(^{19}\) The following passages are particularly illuminating with regard to the motivating effect that the actualisation and institutionalisation of reciprocity is seen to have on the realisation of justice as fairness: “Because we recognize that they wish us well, we care for their well-being in return. Thus we acquire attachments to particular persons and institutions according to how we perceive our good to be affected by them. The basic idea is one of reciprocity, the tendency to answer in kind”. (TJ, 494. Emphases mine). And: “First of all, the unconditional concern of other persons and institutions for our good is far stronger on the contract view. The restrictions contained in the principle of justice guarantee everyone an equal liberty and assure us that our claims will not be neglected or overridden for the sake of a larger sum of benefits, even for the whole society. We have only to keep in mind the various priority rules, and the meaning of the difference principle as rendered by its Kantian interpretation (persons not to be treated as means at all) and its relation to the idea of fraternity... The effect of these aspects of justice as fairness is to *heighten the operation of the reciprocity principle*. As we have noted, a more unconditional caring for our good and a clearer refusal by others to take advantage of accident and happenstance, must strengthen our self-esteem; and this greater good must in turn lead to a closer affiliation with persons and institutions *by way of an answer in kind*”. (TJ, 499. Emphases mine)
to their advantage; that their claims would be adjudicated fairly and that they could pursue their good to the fullest extent once they complied with the principles of justice. In this way agreement is dependent on the fulfilment of the criterion of reciprocity (cf. TJ, 388).

Some further points about the notion of reciprocity are worth noting at this stage. Firstly, it is a recipient-centred principle. This reflects Rawls's understanding of basic justice as rights-centred; we begin with the claims of individuals with divergent conceptions of the good, and justice then involves the mediation of those claims. It is the recipient perspective, as opposed to the agent perspective; it begins with a claim rather than an action.20 From this we reach a second consideration about the nature of reciprocity. As a fundamental principle, reciprocity involves a dual criterion; the "tendency to answer in kind" implies that the principle is only satisfied when both parties fulfil their part of the bargain (TJ, 494). It works as such; just principles are acted upon, and a corresponding just action is returned in kind.

What this also implies is that if the second stage of the criterion fails, the principle is undermined and the entire scheme of justice is called into question. The reason for this is evident when we recall that reciprocity is grounded in and representative of the notion of the person as free and equal, and further that equality is a limiting feature of the original position. Fulfilling the principle of reciprocity is equivalent to expressing oneself as an equal moral person, because this is achieved, according to Rawls, by acting on the principles of justice, which are grounded in reciprocity, or mutual advantage. The principles of justice rely on an understanding of equality as reciprocity, so if the second stage of reciprocity

---

20 As we will see, O'Neill has argued that Kant's ethics begins from the agent perspective that focuses on obligation. This does not imply that rights are subordinate to obligations, rather it emphasises that rights without counterpart obligations are meaningless, and that it is more practical to begin with a focus on action and agency rather than on rights and claims. The agent perspective will be discussed further when we look at O'Neill's specific interpretation of Kant's ethics. For further reading see O. O'Neill, "Kantian Ethics", in Peter Singer (ed.), A Companion to Ethics (Oxford: Blackwell 1993), pp. 175-184; O. O'Neill, "Kant: Rationality as Practical Reason", final version published in The Oxford Handbook of Rationality, ed. Alfred J. Mele and Piers Rawling, (Oxford University Press: Oxford, 2004), pp. 93-109; O. O'Neill, Constructions of Reason: explorations of Kant's practical philosophy, (Cambridge: Cambridge University Press, 1989); O. O'Neill, Bounds of Justice, (Cambridge: Cambridge University Press, 2000).
Practical Reasoning and Transnational Justice

goes unfulfilled, if just action receives no response, the principles of justice themselves are left unsupported. This is because the good will to act in a just way is not methodologically fundamental; rather the claim is prior to the fulfilment of the claim.  

A further implication of beginning from a principle of reciprocity grounded in the recipient perspective is that such a principle must presuppose the legal framework that institutionalises reciprocity as a legal principle. Not only does this mean that, because such a legal framework is presupposed, it stands in need of further and more fundamental justification. This also flags a deep contrast with Kant's account, as he distinguished between the level of legal institutions and the more fundamental level of the moral law, which, unlike legal reciprocity, may have an asymmetrical character in that it is not dependent on a response.

The principle of reciprocity is fundamental to justice as fairness; just action, conceived as fair action, is the manifestation of this principle. This principle guides the construction of the principles in the original position. From Kant, Rawls also takes just action to be equivalent to autonomous action, though, as we will see, he departs from Kant on the crucial question of what constitutes autonomous action. An examination of how the original position models the notions of equality

---

21 As O'Neill has commented: "A conditional willingness to agree if there are terms that will (or would) be agreed on by all binds nobody when the condition does not obtain." See O. O'Neill, "Political Liberalism and Public Reason: A Critical Notice of John Rawls's Political Liberalism", The Philosophical Review, Vol. 106, No. 3 (July, 1997), pp. 411-428, p. 415.

22 O'Neill has argued recently that the justification of the social contract relies on separate claims that move from the more abstract, and fundamental, to the more contextualised, based on actual historical conditions: "Kant's justification of the social contract as an idea of reason relies on three linked claims. Proceeding from the most to the least abstract, these claims are (1) The categorical imperative is the supreme principle of practical reason, so an idea of reason... (2) The universal principle of justice states a version of the categorical imperative, restricted to the public domain... (3) The republican conception of the social contract is a special case of the universal principle of justice, adjusted to certain historical conditions". See Onora O'Neill, "Kant and the Social Contract Tradition", in Kant's Political Theory, ed., Elizabeth Ellis (Pennsylvania: The Pennsylvania State University Press, 2012), pp. 34-35. Rainer Forst has also argued that Rawls must distinguish the universal level of morality from the reciprocal level of law: "moral persons are the authors and addressees of moral norms, citizens are the authors and (as legal persons) addressees of legal norms. As such they lay claim to certain social primary goods, whose specification and distribution must be reciprocally justified in political contexts." See Forst, Contexts of Justice: Political Philosophy Beyond Liberalism and Communitarianism (Berkeley: University of California Press, 2002), p.189.
and freedom will further elucidate how these concepts stand in relation to each other and to autonomous action and the desire to act on principles of justice.

1.1.3. The interpretation of Kantian autonomy in Rawls’s constructivist account of justice

Before I examine Rawls’s interpretation of Kantian autonomy I will comment briefly on the role of freedom in Rawls’s “Kantian” interpretation. An account of how freedom is represented is necessary in order to understand Rawls’s particular account of Kantian autonomy and also therefore later to evaluate alternative interpretations. I hope to show how his reductionist account of positive and negative freedom relates to his understanding of Kantian autonomy. As Rawls is committed to a preference-based theory of action, noumenal freedom to reason about action is reduced to empirical freedom of choice. Yet he must still connect freedom to autonomy, and because he rejected the idea that the agent’s capacity for reason itself could guide the construction of its own principles in abstraction from all desires and preferences, this leaves him with the predicament of showing how certain desires connect to autonomous agency without being considered heteronomous sources of law. Since Rawls identified autonomy with autonomy of the individual or personal self, he had to somehow show how a particular kind of desire could be so essential to the individual self that it was not considered a heteronomous source of law. Rawls’s account diverged from Kant in rejecting an inherent human capacity to be moral, i.e. to reason about and act from principles of justice; he instead distinguished between preferences and “highest-order” desires. His reluctance to describe freedom in terms of innate capacities to reason about action forces him to leverage his understanding of autonomy against the highest-order preferences of individual selves. In contrast to Rawls, in Kant’s account, the autonomy of reason, as opposed to the self, represents an inherent moral capacity for practical reasoning that is neither dependent on desires nor inevitably hindered by natural and social contingencies.
1.1.3.1. Freedom as modelled in the original position

There are two aspects of freedom evident in Rawls's conception, one indicating a form of negative freedom, and one indicating a form of positive freedom.\(^23\) Firstly, in relation to negative freedom, he tells us that the parties can be described for the purposes of the original position as "noumenal selves" (TJ, 255).\(^24\) The parties can be seen as selves outside the causal order, as non-determined selves, because they exist within a framework that has bracketed those aspects of human life that are usually determinate. Actual persons are necessarily under the influence of natural and social contingencies, but the parties are not. Rather, they are completely independent from all normally determinate factors and as such have an empirically defined freedom in the choice of principles. But not only do the parties have freedom in this sense of unrestricted choice, they also have a concern to choose principles for social interaction which they consider to be in their subjective interest. The conditions for the selection are modelled by the structure of the original position; the parties themselves are not inclined toward principles of justice, rather the framework within which they make their choice is designed to bracket information that could lead them toward the selection of unjust principles. The parties therefore are also encouraged, indirectly, to use their freedom in a regulated way, or in a positive sense. This is evident when we consider Rawls's claim that "Kant's main aim is to deepen and to justify Rousseau's idea that liberty is acting in accordance with a law that we give to ourselves" (TJ, 256). Crucially however, in contrast to Kant, it is the design of the original position that represents positive freedom, and not the moral capacities of the agents.

\(^{23}\) In the Kantian sense: negative freedom as freedom from external influence, and positive freedom as a certain way of making use of this freedom. See Onora O'Neill, *Bounds of Justice* (New York: Cambridge University Press, 2000), p. 43. However, whereas according to O'Neill it is the freedom of the will in terms of reasoning about action and justice, for Rawls it is the independence of the self's highest-order desires from external contingencies. O'Neill has objected to this interpretation of Kantian autonomy in several places, most notably in her criticism of appeals to patient autonomy in medical ethics. See O. O'Neill and Neil C. Manson, *Rethinking Informed Consent in Bioethics* (Cambridge: Cambridge University Press: 2007); O. O'Neill, *Autonomy and Trust in Bioethics* (Cambridge: Cambridge University Press: 2002)

\(^{24}\) Already in *A Theory of Justice* it is apparent that Rawls views the Kantian account of freedom as problematically grounded in transcendental idealism, hence already at this stage he attempts to pre-empt these issues by drawing on an empirical conception of freedom and linking it to representative agents and not persons. Indeed, O'Neill has suggested that it is this "reluctance to credit humans with noumenal freedom that lies behind Rawls's reluctance to go the whole way with Kant". See O'Neill, *Bounds of Justice*, p. 45.
themselves, and therefore the agents themselves are not seen as making use of positive freedom. Both negative and positive freedom are represented in the original position by an independence from social and natural contingencies; while the parties' choice is not determined by these factors, it is restricted by their absence. Again this stands in contrast to Kant's account of freedom, where contingencies are no barrier to morality. This brings us to Rawls's account of autonomy, as autonomy, or self-legislation, can be understood as freedom or independence from external sources of law.

1.1.3.2. Autonomy: the "highest-order" desire to act on principles of justice

The purpose of our discussion of Rawls's interpretation of Kantian autonomy will be to further highlight the implications for his account of justice of his early methodological reliance on the hypothetical social contract (i.e. hypothetical consent defined in terms of empirical preference and later justified via actual consent), and his rejection of Kant's view that practical reason could itself constitute the order of moral values (PL, 99). As we have seen, the emphasis on the hypothetical social contract implies that the focus of Rawls's theory from the beginning must relate to the probability of agreement or consent to principles of justice, first as set out in hypothetical terms and then tested against the convictions of actual agents. And the rejection of "constitutive autonomy" implies that he must consider how and why the individual self (as opposed to reason itself), acting from a law it gives itself, counts as autonomous, i.e. in order to explain the value of autonomous action, Rawls must simultaneously show why it is action that is morally valuable and also why it is action that moral agents would want to choose. Given that what agents would want is defined in relation to empirical desires (albeit "highest-order desires"), this implies that his account of autonomous action must contain an account of motivation that links action to desire.

Rawls sums up his interpretation of Kantian autonomy as follows:
Kant held, I believe, that a person is acting autonomously when the principles of his action are chosen by him as the most adequate possible expression of his nature as a free and equal rational being. The principles he acts upon are not adopted because of his social position or natural endowments, or in view of the particular kind of society in which he lives or the specific things that he happens to want. To act on such principles is to act heteronomously. (T], 252. Emphasis mine)

To act from a law that we give ourselves is firstly to avoid action grounded in any heteronomous principles, that is, from any external law. External law in Rawls’s interpretation ranges from social or natural contingencies to most, though not all, of our preferences or desires. Subjective desires and ends are extraneous; principles that express autonomy are those that “apply to us whatever in particular our aims are” (T], 253).

As Rawls understands it, to act from a law that we give ourselves is also to act from a law that expresses our fundamental nature; autonomy, defined as self-legislation, is legislation by the self. For Kant, according to Rawls, this true nature is a conception of ourselves as free and equal rational beings. When we act autonomously, we act on principles that are not conditional upon anything except our basic freedom and equality. Rawls states of the Categorical Imperative that “[t]he validity of the principle does not presuppose that one has a particular desire or aim...[rather it] applies to a person in virtue of his nature as a free and equal rational being” (T], 253). It is this connection between our fundamental freedom and equality and the validity of principles that Rawls attempts to reinforce via the original position.

Rawls outlines the Utilitarian philosopher Henry Sidgwick’s objection to Kantian autonomy, which, in brief, suggests that Kant has not shown why one set of principles selected under conditions of freedom and equality can more adequately express our nature than another, perhaps morally questionable, set; is the scoundrel’s principle not as equally valid an expression of his “characteristic and freely chosen selfhood” as the saint’s? (T], 254-5). Rawls suggests Kant’s response would be to claim that while the free self can choose any “consistent set of principles”, not all such sets express our characteristic freedom and equality; only certain sets of principles express our nature as free and equal. However, he
accepts that Kant has not shown the vital correlation between the moral law and the *expression* of our nature as free and equal rational beings, and he proposes the original position as a remedy to this omission:

My suggestion is that we think of the original position as the point of view from which noumenal selves see the world. The parties qua noumenal selves have complete freedom to choose whatever principles they wish, but they also have a desire to express their nature as rational and equal members of the intelligible realm with precisely this liberty to choose, that is, as beings who can look at the world in this way and express this perspective in their life as members of society. They must decide, then, which principles when consciously followed and acted upon in everyday life will best manifest this freedom in their community, most fully reveal their independence from natural contingencies and social accident (TJ, 255).

The original position, according to Rawls, *models* the conditions of freedom and equality that manifest our true nature, making it possible then to select principles that manifest this nature. In other words, it shows why a noumenal self, defined empirically as simply having complete freedom in their choice, would choose the principles of justice over other principles. It is a “procedural interpretation of Kant’s conception of autonomy and the categorical imperative”; the framework of the original position is designed so as to limit relevant considerations to those that reflect the freedom and equality of persons (cf. TJ, 256). Specifically for Rawls this implies that the “contingencies of nature and society” must not be allowed to influence the decision-making process of the parties; freedom and equality are represented in the original position, via the veil of ignorance, as *independence* from these contingencies (cf. TJ, 265). This, Rawls believes, solves the problem of “expression” in Kant’s account of autonomy. Through the original position the connection between our nature as free and equal rational beings and the principles of right and justice is made apparent; “[o]ur nature as such beings is displayed when we act from the principles we would choose when this nature is *reflected in the conditions* determining the choice” (TJ, 255-6. Emphases mine).

It seems that, on Rawls’s account, persons as free and equal beings are manifestly so when they act in ways that accord with such a conception of the person, and,
conversely, that any action based on principles that deny the freedom and equality of persons obscures and contradicts our true nature.

However, we may follow this interpretation only so long as we can also accept Rawls's claim that the original position is the *most adequate method* for identifying the correlation between the basic freedom and equality of persons and the fundamental principles of right and justice. For Rawls makes two claims: firstly, the above statement that just principles are those that accord with our nature as free and equal, and a second, additional claim, that *these principles are those of justice as fairness*, or principles closely akin to those of justice as fairness. For Rawls, it is not action based on the freedom and equality of persons that expresses the self as free and equal: it is specifically actions based on principles of *fairness* that express the self in this way. He identifies freedom and equality with independence from natural and social contingency, and therefore with fairness and impartiality. This might be a purely academic consideration for Rawls, because he takes these to be equivalent. However, if we do not accept this equivalence, if we question or reject the correlation between the self as a free and equal rational being and the principles of fairness selected via the original position, we may also be forced to reject Rawls' first claim; his account of the principles that reflect freedom and equality entails that they be principles of fairness and is therefore incompatible with any interpretation of freedom and equality that does not arrive at this conclusion.

Furthermore, we can accept this position only if we accept the claim that, for Kant, to act autonomously is to act from a *desire* to act on principles that reveal our true nature, and to act otherwise is to conceal that nature. More specifically, we must accept the conception of autonomy as legislation by the self, understood as the individual self.

Since acting on principles of justice for Rawls does not account for the whole conception of the moral person, the question of the "unity of the self" is raised. Specifically, Rawls considers how his method can argue for coherence between
principles of justice and conceptions of the good. His discussion of the unity of the self allows us further insight into Rawls’s specific Kantian interpretation.

The notion of unity is raised primarily in order for Rawls to argue against teleological or “dominant end” approaches to justice – those that begin with a conception of the good (such as pleasure in hedonistic Utilitarianism) and then seek the unity of the self through maximisation or aggregation of this good (TJ, 562). Right principles, on these approaches, are specified by reference to this dominant end.

Rawls argues that his approach is a complete “reversal of perspective” from dominant end starting points: it begins not with the individual experience aggregated and maximised, but with a “general framework for the whole”, or an outer boundary, within which specific, undefined, individual ends can be freely formulated (TJ, 563-6).

This reversal also introduces an alternative understanding of the “self”; the dominant end is replaced by moral personality as the fundamental aspect of the self (cf. TJ, 563). The unity of the self on Rawls’s account is equivalent to the full realisation of moral personality. “Moral personality” consists of two capacities, a “sense of the good” and a “sense of justice.” These are realised, respectively, through the formulation of a rational plan of life, and “through a regulative desire to act on certain principles of right” (TJ, 561). Moral personality is fully realised when one’s rational plan of life is made to accord with one’s desire to act on principles of justice. Unity of the self then is the “coherence of the plan”; one’s chosen ends are not incompatible with one’s desire to follow principles of right and justice (TJ, 561).

If we recall that it is through acting on the principles of right and justice that we express our true nature, and note that Rawls equates this with the sense of justice, we can say that on Rawls’s Kantian interpretation the unity of the self requires that our conception of the good be consistent with our desire to express our nature as free and equal rational beings. There can be no unity of the self unless
the desire to act on principles of justice is a manifestation of a “fundamental preference” to express one’s nature (TJ, 561).

That Rawls identifies a role for desire in his Kantian interpretation is significant, and the comments he makes in defence of this role in “Kantian Constructivism in Moral Theory” are instructive:

[A] Kantian view does not deny that we act from some desire. What is of moment is the kinds of desires from which we act and how they are ordered; that is, how these desires originate in and are related to the self, and the way their structure and priority are determined by principles of justice connected with the conception of the person we affirm. The mediating conception of the original position enables us to connect certain definite principles with a certain conception of free and equal moral persons. Given this connection, an effective sense of justice, the desire to act from certain principles of justice, is not a desire on the same footing with natural inclination; it is an executive and “highest-order desire to act from certain principles of justice in view of their connection with a conception of the person as free and equal. And that desire is not heteronomous: for whether a desire is heteronomous is settled by its mode of origin and role within the self and by what it is a desire for. In this case the desire is to be a certain kind of person specified by the conception of fully autonomous citizens of a well-ordered society.²⁵ (Emphasis mine)

Rawls begins with the statement that “a Kantian view does not deny that we act from some desire.” He must defend the role of “highest-order” desire as the source of autonomous action because he rejects the possibility that agency can be wholly separated from desires. He endorses instead an empirical conception of agency that assumes the originating source of all action to be some form of desire.²⁶ Given this, his argument can be reconstructed as follows: for Rawls, Kantian autonomy does not imply that all desires need to be heteronomous; given that autonomy is legislation by the self, desires that originate in and are related to the self are properly autonomous. Such desires are specified as “regulative highest-order desires”; they relate to principles of right, which are established as taking precedence over conceptions of the good. In response to the question of how such

a desire to act on just principles originates from the self, the answer is that this connection is facilitated by the original position. The fundamental characteristics of the self, freedom and equality, describe the conditions under which the principles of justice are selected in the original position; the desire to act justly is a desire to act as a free and equal self. The sense of justice then originates directly from within the self; it is internal to the self, and so internal to legislation by the self, and so autonomous.

As we have noted, Rawls rejected the suggestion that reason itself could be autonomous as he identified this claim with problematic or controversial metaphysical assumptions about reason.²⁷ He thought that the claim that reason alone could provide the material for the construction of principles of justice invested too much in practical reason itself without the necessary “fixed ideas”, or later “political conceptions”, drawn from “our” experience.²⁸ However O'Neill has argued that Kant’s claims regarding the autonomy of reason were not reliant on any strong metaphysical claims. Rather, on her account, what is implied by the autonomy of reason is that moral agents’ efforts of communication and reasoning with others are easily thwarted, and unless their reasoning subjects itself to certain conditions, incomprehensibility (in thought and action) is inevitable. These conditions in turn supply the groundwork for the construction of moral principles (and maxims of communication). On O'Neill’s interpretation, it is the moral agent’s capacity for reason that is self-legislated, i.e. free from all external influence, including even highest-order, regulative desires, and legislated only by the conditions of its own possibility.²⁹

²⁷ From Political Liberalism: “Another and deeper meaning of autonomy says that the order of moral and political values must be made, or itself constituted, by the principles and conceptions of practical reason. Let us refer to this as constitutive autonomy. In contrast with rational intuitionism, constitutive autonomy says that the so-called independent order of values does not constitute itself but is constituted by the activity, actual or ideal, of practical (human) reason itself. I believe this, or something like it, is Kant’s view. His constructivism is deeper and goes to the very existence and constitution of the order of values. This is part of his transcendental idealism.” (PL, 99)

²⁸ “A political view, we say, is autonomous if it represents, or displays, the order of political values as based on principles of practical reason in union with the appropriate political conceptions of society and person” (PL 99), and “[w]hat is essential is that justice as fairness uses as basic organizing ideas certain fundamental ideas that are political. Transcendental idealism and other such metaphysical doctrines play no role in their organization and exposition.” (PL, 100)

²⁹ Her position will be elaborated in Part 3. For further reading see O. O’Neill, “Kant: Rationality as Practical Reason” in The Oxford Handbook of Rationality, ed. Alfred J. Mele and Piers Rawling,
1.1.3.3. Congruence: autonomous action as an expression of the good

As we saw from Rawls's comments on the unity of the self, it is vital to his thesis that the principles of right cohere with persons' conceptions of the good. Indeed he viewed the compatibility of the right and the good as essential to the stability of his theory. And not only must the right, understood as the principles of justice as fairness, be compatible with persons' individual conceptions of the good, he also argued further that members of a just society would come to see the principles of justice as a fundamental and regulative component of their conception of the good. This he termed the "congruence" of the right and the good.

The sense of justice is fundamentally a desire to act on principles of justice because such action is an expression of the self as a free and equal rational being. Add to this the Aristotelian Principle, that "human beings enjoy the exercise of their realized capacities... and this enjoyment increases the more the capacity is realized" (TJ, 426), a principle described by Rawls as "a principle of motivation" (TJ, 427), and we have the claim that persons in a well-ordered society have a highest-order desire to realise their true nature by following principles that reflect that nature, and that this expression is also "a fundamental element of their good" (TJ, 445). This claim is central to Rawls' argument for the congruence of the right and the good in *A Theory of Justice*, his attempt to show that "being a good person is indeed a good" (TJ, 398).

Though Rawls's congruence argument is not entirely based on the Kantian and Aristotelian principles, they are at this point decisive. Without the Kantian element the argument claims that "participating in the life of a well-ordered society is a great good" (TJ, 571). Cooperation is essential not only for the benefit of society as a whole but for the good of each individual also: members cannot realise fully their "latent powers" without the "cooperative endeavours of others" (TJ, 571). However, if we value cooperation, we must also value "the principles of its regulative conception"; a scheme of mutual advantage can only succeed if the rules

and principles that make it possible are both affirmed and observed. If we see cooperation as a good, we must also see its enabling principles of justice as a good.

The Kantian element in the broad sense then involves viewing the principles of justice as the expression of our true nature. Persons not only affirm such principles because they make the scheme of cooperation possible, but because "acting justly is something we want to do as free and equal rational beings" (TJ, 572. Emphasis mine). Congruence is achieved in this way because persons believe the expression of their true nature to be a fundamental element of their good. The right is not only a good in itself, but it is a good that regulates all other conceptions of the good. Rawls thinks this understanding is Kantian because he links Kantian moral autonomy with the desire to act justly.

The non-Kantian argument for congruence Rawls employs is derivative; persons view the principles of justice as good because they are essential to a cooperative scheme that is valued and viewed as a good. The congruence argument based on Rawls's Kantian interpretation is more direct. The sense of justice and the expression of one's nature are "both dispositions to act from the same principles: namely, those that would be chosen in the original position" (TJ, 572). Here the principles of justice are viewed as a good because they are equivalent to a fundamental element of each person's good. As we have seen, expressing one's nature by acting on principles of justice is how Rawls conceives of Kantian autonomy. Therefore this argument for congruence makes the claim that all members of a well-ordered society could, or would, come to endorse a conception of moral autonomy that views the desire to act on just principles as equivalent to the desire to act as a free and equal person.

This appeal to congruence is not trivial: Rawls leverages much of the justification for justice as fairness on the claim that moral persons, with due reflection, may indeed endorse the account of justice linking principles of right, higher-order interests and moral autonomy to subjective interests and conceptions of the good.
1.1.4. Justifications: considered convictions and reflective equilibrium

Despite the proposed effectiveness of the original position as a device of representation that could generate principles of justice for a bounded society of moral persons, Rawls does not make this procedure methodologically fundamental; he does not claim it is the only source of justification for his theory of justice. Rather throughout his writing he also consistently appeals to the existing principles, sentiments and views of the audience to whom he directs justification.\(^\text{30}\)

The agreement reached under the original position can be understood purely hypothetically. Therefore, Rawls suggests, the principles arrived at risk being irrelevant to actual persons unless they can be shown to reflect the judgments of real persons in some other way:

"We shall want to say that certain principles of justice are justified because they would be agreed to in an initial situation of equality. I have emphasised that this original position is purely hypothetical. It is natural to ask why, if this agreement is never actually entered into, we should take any interest in these principles, moral or otherwise. The answer is that the conditions embodied in the description of the original position are ones that we do in fact accept. Or if we do not, then perhaps we can be persuaded to do so by philosophical reflection. (T], 21. Emphases mine)

The guiding suggestion is that Rawls's principles of justice should cohere with "our" "considered convictions"; pre-established judgments belonging to the members of the bounded society should, upon reflection, exhibit a measure of

\(^{30}\) O'Neill has commented on the pivotal role of Reflective Equilibrium with regard to justification: "Although actual agreements and contracts have normative force, hypothetical ones do not, and unless Rawls provides further reasons for thinking that principles chosen in a hypothetical fair situation are appropriate in the far-from-fair situations in which we find ourselves, the justification is incomplete." O'Neill, "The Method of A Theory of Justice", p. 38. Forst, however, takes the "freestanding" justification, relating to those aspects of Rawls's justificatory strategy not reliant on considered convictions (and later overlapping consensus), to be more fundamental. He argues that those aspects that relate to explaining the "possibility of social stability" are "subordinate" to the level of "freestanding" moral justification". Forst, Contexts of Justice, p. 174. However, based on O'Neill's critique, I suggest that this concern for "social stability" is implicitly integral to all aspects of Rawls's justificatory strategy, because he takes his starting point to be a hypothetical agreement between moral agents that asks which principles such agents would want to agree to, and not the more basic question of which principles they could agree to. His starting point implies that at some point Rawls must show why actual agents would want to agree to such principles, and therefore even his "freestanding" justification is from the beginning tied to questions of preference and stability.
conformity with the principles of the original position. It is this measure of coherence or conformity that he terms "Reflective Equilibrium".

Rawls is careful to attempt to avoid the charge that his justification falls back on "actual" judgments of persons, whatever they might happen to be. He therefore sets out some guidelines or criteria for distinguishing "considered" from unconsidered judgments. He outlines his criteria as follows:

[T]hey enter as those judgments in which our moral capacities are most likely to be displayed without distortion... Considered judgments are simply those rendered under conditions favourable to the exercise of the sense of justice, and therefore in circumstances where the more common excuses and explanations for making a mistake do not obtain. (T], 47-8)

It is apparent that Rawls conceives of the criteria for considered judgment to be similar to those that underpin the construction of the original position, specifically the bracketing of considerations deemed irrelevant to justice. Indeed the entire construction of the original position is designed to reveal our "moral capacities" without "distortion", just as Rawls suggests considered judgments do. Rawls also comments, with regard to "considered convictions", that "[w]e think we have examined these things with care and have reached what we believe is an impartial judgment not likely to be distorted by an excessive attention to our own interests" (T], 19-20. Emphasis mine)."[E]xcessive attention to our own interests" amounts to precisely the contingencies Rawls sought to avoid in his meticulous moulding of the original position.

An apparent objection at this point relates to the seeming disparity between the two modes of justification that are to "cohere" within reflective equilibrium. While the original position is a constructive procedure for selecting principles of justice, considered convictions appear to be just those that are given in a particular culture or society. The obvious corollary then is that, because such convictions appear as unvindicated premises, their status as the foundation of a justificatory strategy is drawn into question.
I would suggest that while it is not incorrect to object to the status of considered convictions in Rawls's justificatory strategy, the appropriate objection is not that Rawls assumes the unvindicated convictions of members of a particular culture or society to be the unquestionable basis for judgments of justice. To understand why, we must consider why Rawls thinks it uncontroversial to present the convictions of moral persons, qua citizens, as such a basis.

I will briefly present Rainer Forst's defence of Rawls, discussed in his early work, *Contexts of Justice*, against this type of objection. I will next suggest that while I agree with Forst that it is a misunderstanding to suggest that considered convictions in Rawls's account are presented as a basis for justification without an attempt to vindicate their status as such, this does not absolve Rawls of the charge that the justificatory role of considered convictions is grounded in unvindicated premises. I argue that the objection applies at a deeper, foundational theory level, to his interpretation or understanding of the principles of practical reason. To support this I will also note some comments from Rawls that stress the role of reflective equilibrium as highlighting the origins of both considered convictions and the principles of justice in "common sense" or "shared" ideas.

Though Forst's immanent critique of Rawls's theory of justice refers mostly to *Political Liberalism*, it focuses on broader and deeper theoretical aspects of Rawls's account, broad: the potential tension between the universalist and contextualist interpretations of the theory, and deep: the level of shared principles of practical reason. Therefore I include his comments at this early stage, as a discussion of how considered convictions relate to practical reason in *A Theory of Justice* is crucial to understanding the role of the ideals of citizenship and liberal democracy in *Political Liberalism*. This discussion will facilitate an explication of the seeming arbitrary nature of the imposition of ideals of citizenship and liberal democracy and hence identify and trace the fundamental theory decisions that lead Rawls to his particular understanding of public justification and public reason in his later works.

---

31 See Rainer Forst, *Contexts of Justice*, pp. 154-229. Further references in text (CJ)
Forst first suggests that although there appears to be a contextualist element to Rawls's strategy of justification, this relates to social stability and is subordinate to a less context-sensitive "freestanding" justification. Next, he addresses the suggestion that this "freestanding" justification is still only a reflection of a contextualised or particular self-understanding. He rejects this suggestion on the basis that it is "incompatible with Rawls's claim to justify a "reasonable" conception that—unlike conceptions of the good that question the priority of justice... puts forward stronger reasons than the reference to "our" practices allows" (CJ, 174). Rawls's emphasis on reasonableness mitigates against him adopting controversial premises grounded in the shared thinking of particular societies.

He then addresses the objection that Rawls's theory is ultimately justified by a "fundamental commitment to the liberal political ideal" so that "[w]hat forms the normative starting point is not therefore the mere presence of the liberal ideal in a particular political culture but the normative ideal itself—an ideal, nevertheless, that can in turn be justified only as a "comprehensive" liberal doctrine of the good" (CJ, 174-175): the reasonableness of the theory is called into question because it ultimately relies on ideals and conceptions that belong to a liberal account of the good that cannot be justified in the political sense that Rawls aims for.

It is at this point in particular that Forst's comments are of note. His defence of Rawls here highlights the important chain of reasoning that connects the principles and ideas of practical reason to the considered convictions, or reasoning, of citizens, and then to the ideals of citizenship, hence highlighting the universalist assumptions Rawls makes regarding what initially appear as contextualist claims.

Forst argues, regarding the suggestion that Rawls's theory is tied to a liberal doctrine of the good, that Rawls "contradicts this interpretation by emphasising that his fundamental idea of social cooperation and the attendant ideas of free and equal persons and of well-ordered society... are "ideas of practical reason.... that 'cannot be reasonably rejected' in a moral sense" (CJ, 174-175. Emphasis mine).
Practical Reasoning and Transnational Justice

What this implies is that ultimately, according to Forst, the contextualist objection is avoided so long as Rawls's theory of justice "rests on practical reason". He writes:

For that reason, the conception of "justice as fairness" begins not with contingent "shared understandings" because they are contained in a particular political culture, but with conceptions of person and social cooperation that must be contained in such a culture—and indeed necessarily so if the culture raises the claim to being a democratic one that rests on a shareable, reasonable foundation. Without these conceptions of practical reason there is no democratic, legitimate society. They are inherent in the fundamental principle of public justification: a just and publicly justified basic structure of society—a structure that expresses citizens’ "shared and public political reason" (1993a, 9) —must rest on these conceptions since they themselves are part of the idea of public reason. (CJ, 175)

Furthermore, he adds:

That the theory of justice "starts" (14) with certain fundamental ideas of a democratic political culture is therefore justified in the "philosophical background of political liberalism in practical reason" (xiv), not in a more or less conventionalist alignment of the theory. That we can reconstruct the right from familiar concepts does not mean that it is right because it corresponds to "our" familiar conceptions. (CJ, 175)

The contextualist interpretation is inappropriate because, Forst claims, Rawls grounds his theory in universally shareable principles and ideas of practical reason. The notions of impartiality, reciprocity and social cooperation that constitute the conceptions of the "reasonable" person (first as moral person and later as citizen) and the well-ordered society are "latent in common sense". Forst describes this appeal to universal practical reason as a "Kantian interpretation of the new shape of Rawls's theory" (CJ, 175). Rawls's account is not grounded in an appeal to the particular convictions of members of certain societies because at heart it is trying "to reformulate the Kantian principles of moral autonomy as acting according to universally justified principles with a view to drafting principles of justice for a

---

32 Furthermore, Forst argues that "[t]he basic ideals of the moral person and the well-ordered society are ‘available to the common sense of any thoughtful and reflective person’... That the ‘fundamental ideals’ to which he [Rawls] refers are immanent in a democratic political culture does not mean that their claim to validity is restricted a priori to this culture". (CJ, 182)
basic structure of society" (CJ, 176. Emphasis mine). Rawls's account is not contextual because it is seeks to offer universal-moral justification. Forst interprets the criteria that constitute the principles of practical reason as the demand that "general norms must be generally justified" in a process he describes as "recursive" and "discursive" (CJ, 176). In this way the considered convictions of moral persons in reflective equilibrium with the principles of justice, and later the public justification of principles in a liberal democratic society, cannot be understood as contextualised because they rely on an intersubjective process of universal justification that is grounded in a form of practical reason the principles and ideas of which "cannot be reasonably rejected' in a moral sense" (CJ, 175).

In this way the apparently shared form of reasoning that underpins both considered convictions and the procedures of the original position become apparent. According to Forst, the method of reflective equilibrium requires that

[T]o find out what principles of justice can raise a claim to validity, they must be acceptable as just principles to free and equal persons—and the conditions for this free and equal acceptance must be mutually clarified and laid down in the equilibrium between the specific resulting principles and "our" moral judgments."... Here, the original position takes up a mediating position: it is the "rationalizing" medium on the basis of which generally acceptable judgments on fairness and impartiality can be formulated in such a way that substantive principles of justice spring from it. From this method of bringing intuitions, principles, and abstract procedural conditions coherently together, Rawls expects the possibility of sense of justice about itself in a reconstructive manner. (CJ, 179. Emphases mine)

Through reflective equilibrium, the conditions and considerations that intuitively guide and underpin "our" considered convictions are brought to light in the constructive procedure. The principles and ideas of practical reason that all "reasonable" persons implicitly accept are formalised and proceduralised, and in doing so the causal relationship between such principles and both "our" considered convictions and the principles of justice as fairness is exposed and brought into view for comparison and evaluation. The causal influences are not equivalent: on Rawls's account the original position is a necessary device because in everyday reasoning persons struggle to abstract themselves from the social and natural contingencies that bias them against arriving at appropriately just
principles. However, persons do have some sense of how the principles of practical reason ought influence their judgments, and in reflective equilibrium what is made apparent is how the original position clarifies and isolates this influence and hence yields the appropriate principles of justice.\footnote{33 "When a person is presented with an intuitively appealing account of his sense of justice (one, say, which embodies various reasonable and natural presumptions), he may well revise his judgments to conform to its principles even though the theory does not fit his existing judgments exactly. He is especially likely to do this if he can find an explanation for the deviations which undermine his initial confidence in his original judgments and if the conception presented yields a judgment which he finds he can now accept. From the standpoint of moral philosophy, the best account of a person’s sense of justice is not the one which fits his judgments prior to his examining any conception of justice but rather the one which matches his judgments in reflective equilibrium. As we have seen, this state is one reached after a person has weighed various proposed conceptions and he has either revised his judgments to accord with one of them or held fast to his initial convictions (and the corresponding conception)". (TJ, 48)}

Despite this defence of Rawls’s universalist framework, Forst concludes that certain theory decisions undermine the universalist-moral aspirations of Rawls’s account, in particular what he describes as the “political-‘substantive’ loading of the conception of moral person in regard to the theory of primary goods” (CJ, 176).\footnote{34 The “ideal initial situation connects rationality assumptions and empirical considerations concerning necessary primary goods with a Kantian conception of practical reason and of the impartiality and autonomy of principles that, without the influence of individual or social contingencies, apply to all ‘reasonable and rational persons’... The procedural principle of general and autonomous justification is connected, in a hypothetical thought experiment, to certain assumptions about a ‘rational’ choice of subjectively desirable primary goods”. (CJ, 180)} Such a loading of the conception implies much stronger assumptions than a purely procedural approach. According to Forst these strong assumptions “make Rawls’s theory weaker in respect of normative justification” (CJ, 188). In the end he thinks Rawls fails to offer a conception of the moral person that can be defended as a universal, context-independent idea of practical reason. Rather Rawls’s context-dependent assumptions cannot raise the claim to being as “reasonably” justified as the basic conceptions of practical reason. He is faced with a dilemma: either his assumptions are reason-based in a strict sense or they contain substantive conceptions of social membership; on both cases the theory is “political”, however, in a more or less context-bound sense. Thus “moral person” is on the one hand explicates on the basis of moral concepts and, on the other, interpreted with reference to the implications of “citizenship” in a more concrete meaning—for instance, with reference to “citizens’ needs. (CJ, 188)
Rawls's conceptions of person and society are not merely uncontroversial representations of the principles and ideas of practical reason: they contain substantive criteria and are therefore not as amenable to universal justification in a "discursive" and "recursive" procedure of rational argumentation as Rawls assumes.

While the solution for Forst is that Rawls must more clearly distinguish the moral person from the person as citizen and separate moral-universal from legal-particular contexts (cf. C], 189), what is interesting for our purposes is how this illustrates the level at which the contextualist objection must be situated. I would suggest that the contextualist issues arise not only from a loading of the conception of the moral person with content relating to "citizens' needs", but from Rawls's understanding of Kantian practical reason itself. In contrast to Forst, I do not believe that Rawls's account can be rescued by distinguishing the moral person from citizen, and universal from legal-political context of justification. Rather the issue is that the basic criteria Rawls takes to constitute the principles of practical reason are not an appropriate basis for claiming universal-moral validity. Forst is correct that to accuse Rawls of contextualising his account simply because he brings in particular conceptions of person and society is a misunderstanding of his aims: such a position neglects to consider that Rawls takes such conceptions to be universally justifiable and believes them to be representations of principles that are neither indigenous to nor confined to liberal democratic societies, but merely immanent in, or represented in, such societies. However, the contextualist objection remains valid, only now at a deeper theoretical level: by loading practical reason itself with substantive content regarding the formal criteria that are to influence the decision-making procedures of moral agents, Rawls introduces a contextual element that is incommensurable with his moral-universalist aspirations.

I would go further than Forst in suggesting that identifying conditions of impartiality, fairness, reciprocity etc. as forming the principles of practical reason and the content of reasonableness does indeed tie the account to a specific
articulation of society and the person as represented by citizens and liberal-democracies: it goes beyond what can be legitimated in a universal-moral context as “unreasonable” to reject. At the very least it puts forward substantive content that stands in need of universal justification and cannot therefore form the criteria for universal justification.

I also suggest that this relates to Rawls’s understanding of justification. On Rawls’s view justification and legitimacy are tied first to the hypothetical social contract, or hypothetical agreement, and next to the actual intersubjective endorsement or acceptance of such an agreement. As will become increasingly apparent later when we discuss Political Liberalism, the focus on normative hypothetical agreement justified through actual endorsement forces Rawls to seek out shared content from which to ground the likelihood of further agreement. Hence at this early stage he must define substantive conceptions of person and society that can contribute toward a framework for determining which principles have a probability of acceptance by moral persons.

What our discussion of the tension between the universalist and contextual aspects of Rawls’s theory (that first becomes apparent in A Theory of Justice, where considered convictions are presented as supporting the normative claims of the hypothetical social contract) has hopefully illuminated is the connection between Rawls’s decision to link justification with a hypothetical social contract, which he attempts to justify through actual agreement, and context-dependency of his account. I would suggest that this primary focus on the likelihood of the acceptance of principles of justice leads into theoretical terrain where considerations of what persons already agree to become relevant, and this, if it is to be anything more than the lowest common denominator of ethical doctrines, can easily lead to contextualised or relativised claims about reason and agency. This is because, as Rawls has emphasised several times, for further agreement on fundamental issues to be likely we must begin from some shared basis (PL, 299). While this aspect of his theory is emphasised most in Political Liberalism with its account of public reason, it is apparent at an early stage that the hypothetical agreement in the social contract is prompted by a prior agreement or acceptance.
of certain shared principles of practical reason. However, to propose supposedly “shared” premises with substantial content relating to conceptions of person and society is not equivalent to the claim that principles of justice can be constructed from abstract premises. Alternative accounts that do not begin by seeking the conditions for probable agreement could avoid the problem of trying to identify uncontroversial, universally acceptable substantive content for an account of practical reason.  

In summary, the context-dependent element to Rawls's justificatory strategy is grounded at a fundamental theoretical level in an account of practical reason that views substantive principles of practical reason (equality, reciprocity, impartiality), and their corresponding representations in particular conceptions of person and society, as necessary materials for a constructivist project. The explanation for this assumption lies in Rawls's early methodological decision to begin with the hypothetical agreement of agents in an initial situation of fairness. The framework of the original position is designed to demonstrate why such agents would choose certain principles, hence the bracketing of contingencies and the reliance on an account of rational choice borrowed from economic theory. And given the hypothetical nature of the construction, it then becomes necessary, if the theory is to have any relevance for actual persons, to show why actual moral agents would also endorse these principles. Therefore from the beginning the focus of the theory is on the likelihood of agreement on principles of justice. Hence the question of stability is present from the beginning, and permeates not just the overtly contextual elements of the theory but also those elements that suggest a universal-moral justification. And, given Rawls's assertion that agreement is only likely or probable where it is predicated on agreement (shared premises), it is  

According to O'Neill it is possible to do without loading practical reason with potentially controversial or overly context-dependent substantive content. Instead she suggests returning to what Rawls rejected, namely Kant's suggestion that practical reason itself could provide the material for the construction of just principles, without the use of particular specifications of persons or societies. She argues that it is a misunderstanding to view Kant's claims about the capacity of reason to provide its own principles as resting on controversial metaphysical assumptions, rather the claim relates to the conditions for the possibility of communication and agreement between free moral agents. Hence the focus is brought back to possible agreement, with probable agreement left aside. This will be explicated in detail in Part 3 where we will discuss her alternative Kantian account, and her critique of Rawlsian public reason.
necessary to consider what the content of such shared premises might be, and hence the danger of slipping into relativised or contextualised accounts of reasoning arises.

This central concern to explain the probability of agreement, i.e. stability, becomes even more pressing in Rawls’s later work, culminating in his revised account of liberal justice in *Political Liberalism*, specifically with the introduction of the “overlapping consensus”. Having traced the roots of the contextualist objection to his account of justice back to his account of practical reason and to his early decision theories relating to the hypothetical social contract, it will be possible to identify the “overlapping consensus” as a context-dependent consensus that is incapable of forming a basis for practical reasoning about justice for actual modern liberal democratic societies, either as a form of reasoning internal to such societies or as a framework for reasoning about international justice. Such a consensus will be shown to be inappropriate to the modern context of justice because it relies on an understanding of practical reason that includes *unvindicated and contextualised* criteria for what is to qualify as “reasonable” in terms of public reason and public discourse.

### 1.2. Political not comprehensive liberalism

The analysis of *Political Liberalism* begins by noting how Rawls had, by this point, dropped the explicitly Kantian element to his justification. He retained what he saw as the less problematic strategy, which relied on the endorsement of the persons to whom his account of justice is directed, now defined in specific terms as citizens of a liberal democratic polity (1.2.1.). In response to his growing concerns regarding the likelihood of reaching consensus over a conception of justice in a liberal democracy that fostered a variety of “reasonable” comprehensive doctrines, Rawls opted to remove from his account of justice any claims relating to a congruence between the right and the good. Now he sought to re-articulate and emphasise those aspects of his justificatory strategy that focused on the likelihood of achieving a consensus on justice within the boundaries of a liberal democratic
society. In order to achieve this he needed to re-articulate his basic concepts in a way that he believed would be acceptable to a plurality of persons whose only point of convergence on matters of ethics and justice was a shared ideal of citizenship, and hence a capacity for reason that could be described as "citizen's reasoning". He therefore presented fresh accounts of freedom, equality and autonomy that were now designed to represent the necessary attributes of citizenship (1.2.2.). Freedom was now grounded in a capacity to form and revise a conception of the good, while equality denoted the minimum threshold of participation in a system of social cooperation. The vindication offered for these conceptions is now that they represent how citizens see themselves.

It should be noted that these new accounts were not radically different to his earlier accounts of freedom, equality and autonomy. The major difference now was the dropped reference to the "free and equal self", which no longer plays a role in his theory. It could also, however, be argued that his earlier accounts of freedom, equality and autonomy, given their empirical characterisation, already bore little resemblance to Kant's conceptions.

Freedom and equality now also respectively delineate the meaning of "rational" and "full" autonomy. Expressing one's free and equal nature as being part of one's "good" is now replaced with a purely formal understanding of the good; autonomy now involves expressing freedom and equality in practical political terms as capacities to participate in society and pursue conceptions of the good. Rational autonomy is a freedom to pursue subjective ends and full autonomy involves an objective limit to rational autonomy based on the formal, politically defined equality of all citizens.

With these conceptions in place, Rawls could now present what he defined as a "freestanding" conception of justice, i.e. one entirely detached from conceptions of the good and reliant only on ideas that were limited to public political domain (1.2.3.). This space could hopefully offer a neutral, mediating ground between conceptions where important matters of basic justice could be decided; starting with the justification of principles of justice and extending to a framework for
working out the application of political principles and ideals to the basic structure, i.e. in the specification of basic rights and liberties and the resolution of disagreements on such matters. Since the aim was to seek consensus on political questions, again Rawls's focus was primarily on the factors that influenced the likelihood of agreement in the political sphere. This implied that ethical questions ought to be bracketed or avoided entirely, if possible, given their propensity toward division and disagreement. It also, and crucially, implied for Rawls that a shared core of agreed content must be sought out, around which a consensus could form. It is for this reason that he argued for a substantive account of political values and ideals to underpin the political conception of justice.

Finally, I examine Rawls's revised justificatory strategies (1.2.4.). The analysis examines first, in relation to pro tanto justification, how Rawls is able to move from the "moral person" to the "citizen" while also claiming that his justification is not context-dependent; the ideal of citizen is for a Rawls a necessary accompaniment to practical reason, which retains for him universal-moral character. As we have noted, he rejects Kant's assertion that practical reason can provide its own standpoint for reasoning about justice, rather it must be a view "from somewhere". He argues that this must be the standpoint of the citizen, who he takes to be the embodiment of the principles of practical reason. Rawls equates practical reason with the point of view of citizens, and this is how citizens' reasoning can become the basis for an overlapping consensus in a liberal democracy populated by a diversity of ethical doctrines. In finishing my analysis of this first stage of justification I ask how Rawls reconciles the shared reasoning of citizens' with their designation as holders of diverse and apparently irreconcilable points of view. I suggest that Rawls has not offered clarity on whether citizens' reasoning is singular (shared) or plural (diverse).

The next stage of justification is "full" justification and reflects the process of reflective equilibrium, now described as a process whereby individual citizens embed the principles of justice as a "module" within their respective ethical doctrines. This is the first stage toward the overlapping consensus, though it is not complete yet because citizens only reason from their own perspective; it has not
yet been seen whether the principles can be intersubjectively or collectively endorsed in the public forum.

At the third stage of justification, “public” justification, citizens gain a formal awareness of others conceptions of the good; they must now see whether the conception of justice they endorse is also accepted by other “reasonable” comprehensive doctrines, though they need not take account of the actual content of other doctrines. Though it could be argued that the earlier stage of “freestanding” justification is central, I suggest that it is this level of “public” justification that is crucial to Rawls, or at the very least it is in no way subordinate to the earlier levels. This is apparent from Rawls’s concern for the “stability” generated by the political conception of justice. I also emphasise that this concern has been apparent from the very first methodological decisions in A Theory of Justice. I also examine why Rawls links stability to legitimacy in his account of justification. Whereas other accounts differentiate between stability (i.e. actual endorsement and consensus) and democratic legitimacy (where principles are justified to all), Rawls sees the two as interlinked and inseparable. Again, I argue that this can be traced back to his understanding of the social contract as first hypothetical and then actual, and of practical reason as not constituting its own authority. These factors imply that Rawls does not distinguish between possible consent and probable consent. This can be most clearly distinguished from O’Neill’s account of Kant’s ethics, which differentiates the level of principled justification (i.e. possible consent) from an empirical calculus of probability.

The corollaries of these considerations are, firstly, that Rawls must argue for a substantive (as opposed to formal or procedural) account of political principles to underpin political reasoning in a liberal democratic society, and, secondly, that a consensus around such political principles is necessary as the foundation of a “public basis of justification”. Public justification is linked to practical reason via the objective point of view; it is represented by the point of view of citizens, which is grounded in practical reason and which Rawls takes to be the measure of objectivity. The final section addresses how this shared basis of political reason is
Practical Reasoning and Transnational Justice

proposed by Rawls as the basis of “public reason” in liberal democratic societies (1.2.5.).

1.2.1. Revised starting points and amended justificatory strategies

Political Liberalism saw quite a radical reconstrual of these initial premises and arguments. What motivated this reassessment was Rawls’s increasing discomfort with what he perceived to be an over-reliance within his theory on supposedly controversial assumptions. In the intervening 20+ years between *A Theory of Justice* and *Political Liberalism* Rawls became more acutely aware of and concerned by the ever-increasing plurality of ethical discourses and worldviews that constituted modern liberal societies. Faced with conflict between the many competing conceptions of the good, he began to have serious doubts about whether his account of justice could ever hope to achieve *stability* within such a plurality, given that stability implied endorsement by the people of the fundamentals of a conception of justice. His original account demanded endorsement of both a congruence between the right and the good, and a procedure for considering “our” convictions, each of which relied on a Kantian notion of the person as a moral being whose fundamental nature was expressed by principles of justice. A corollary of such a demand is that such an account would be defeated if some “reasonable” members of a given liberal society rejected the notion that our fundamental nature was that of free and equal persons, and correspondingly that principles of justice reflected this nature and so expressed also an aspect of “our” “good”. Given the deep diversity of opposing conceptions of the good in liberal societies, such a defeat seemed inevitable. Rawls began to believe that the demands of *A Theory of Justice* were too great.

This was not, however, merely a defeatist concession to a relativist account of conceptions of the good that offered little hope for distinguishing the reasonable from the unreasonable. Rawls was not claiming that this “fact” of ethical pluralism meant we were helpless in the face of odious and/or ungrounded worldviews.
Rather he posited what he termed “the fact of reasonable pluralism”. (PL, 55. Emphasis mine)

Rawls did not see the source of ethical pluralism to lie solely within a conflict between reasonable and unreasonable; he asserted that “reasonable disagreement” could and did also exist. To explain this he appealed to what he called “the burdens of judgment” (PL, 55). This entailed that “reasoned” thought or debate, or “the correct (and conscientious) exercise of our powers of reason and judgment”, inevitably resulted in a multiplicity of reasoned and reasonable perspectives. Put another way, such procedures certainly could not be relied on to produce reasoned agreement. The “burdens of judgment” dictated that there was no possibility of reasonable agreement on those matters that concerned the content of what he now called reasonable “comprehensive doctrines”. The free exchange of reasons between a variety of such doctrines could always and inevitably be expected to end in “reasonable disagreement”. Therefore, no such doctrine could be expected to “serve as the basis of lasting and reasoned political agreement” (PL, 58).

Note the distinction between, on the one hand, the claim that dialogue between ethical traditions frequently ends in discord and, on the other hand, the claim that such outcomes are inevitable. Rawls uses the former to infer the latter, however the latter (Rawls’s conclusion regarding “irreconcilable conflict”) is an unvindicated premise that is based on the experiences of religious traditions in Europe in past centuries, in particular on the Peace of Westphalia, and on the “burdens of judgment”. It is not readily apparent, given the specificities of historical, political and social contexts, why a modus vivendi agreed to by the continental republics of the 17th century should serve as a premise for suggesting that all future efforts at exchange and dialogue between modern ethical communities are doomed toward at best an agreement to disagree. And while the “burdens of judgments” confirm that our capacity to reason leads us in many different directions, it also does not serve as a self-evident premise for the assertion that agreement on divisive issues is eternally out of reach. Rawls himself has suggested that we may “revise” our considered judgments, comprehensive
doctrines and rational plans of life in the light of better reasons. In *A Theory of Justice* he suggests that “[w]hen a person is presented with an intuitively appealing account of his sense of justice (one, say, that embodies various reasonable and natural presumptions), he may well revise his judgments to conform to its principles” (TJ, 48). And with regard to the responsibilities of citizens he suggests that “given their capacity to assume responsibility for their ends, we do not view citizens as passive carriers of desires. That capacity is part of the moral power to form, revise, and rationally to pursue a conception of the good” (PL, 186).

Granted, one could object that only arguments grounded in supposedly noncontroversial abstract claims could be candidates for such revisions of deeply held commitments, and that citizens only revise their ends in relation to their expected share of primary goods (PL, 186).

However ethical doctrines are not always (or perhaps even usually) monolithic units whose acceptance must be *total*. Rather they may be fragmentary, with different aspects and dimensions that are accepted or endorsed to varying degrees. The claims, values and principles that constitute such “doctrines” may be interlocking or separate, mutually supporting, removable or revisable. It cannot be ruled out in advance that reasons that have the capacity to count as abstract, acceptable reasons on which to base shared principles or values might be found within ethical doctrines, nor that these will be limited to the principles of impartiality and fairness as reciprocity that underpin the political conception.36

It is also worth asking why, on the one hand, comprehensive doctrines are taken as immutable in terms of their ethical content, yet on the other hand, citizens are

---

36 Jürgen Habermas has suggested that the claims that comprise comprehensive doctrines must be analysed separately: “Because such doctrines are "comprehensive" in precisely the sense that they offer interpretations of the world as a whole, they cannot merely be understood as an ordered set of statements of fact; their contents cannot be expressed completely in sentences that admit of truth and they do not form a symbolic system that can be true or false as such... under these premises it makes sense to analyze the different validity claims that we associate, respectively, with descriptive, evaluative, and normative statements (of various kinds) independently of the characteristic complex of validity claims that are obscurely fused together in religious and metaphysical interpretations of reality. See Jürgen Habermas, "Reconciliation Through the Public use of Reason: Remarks on John Rawls's Political Liberalism", *The Journal of Philosophy*, Vol. 92, No. 3. (March, 1995), pp. 109-31, p. 126.
expected to be able to revise their convictions or doctrines when presented with "reasonable" arguments. It seems to suggest a double-nature for citizens: one rigidly attached to the entirety of a particular ethical worldview, the other reasonable and adaptable to the concerns and viewpoints of others. And how does the appeal to "irreconcilable conflict" cohere with Rawls's assertion that "[c]itizen's learn and profit from conflict and argument, and when their arguments follow public reason, they instruct and deepen society's public culture" (PL, lv).37

Putting aside these considerations, this implied for Rawls that it was hopeless to expect a stable consensus38 to grow around any account of basic justice that was perceived to rely on overly demanding conceptions of persons or societies. Consensus must instead be sought by removing or eliminating all such controversial assumptions, and attempt must be made to offer strategies of justification that would potentially be accepted by a plurality of "reasonable" worldviews.

Rawls did not seek to completely alter the content of his account of justice. Rather he sought to detach his starting points from what might be perceived as reliance on unacceptable premises, and instead to argue, first, that his were "freestanding" conceptions that could be integrated into a multiplicity of comprehensive doctrines, and, second, to amend his strategies of justification: rather than relying on persons' to reflectively endorse a specific account of justice from both the perspective of the right and the good, Rawls sought to ground his justificatory

---

37 The obvious response is that conflict in exchange and dialogue is only positive when it relates to the exchange of public political reasons, because only these satisfy the criterion of reciprocity. However, as I will suggest, the actual work of public discourse that occurs in Rawlsian public reason is limited, with much work transferred onto the duties of the citizen before they enter the public sphere, and much of the rest allocated to an algorithmic process for identifying answers based on the principles and values of the political conception of justice. There may in fact be very little left to "conflict" or "argue" over, in which case the value from such discursive public reasoning seems to be lost in Rawls's account of public reason. Also, it is not certain that ethical doctrines cannot articulate their own visions of justice that secure the capacities (and rights) of individual agents (fears for which ground the demand for reciprocity). It may be possible for ethical doctrines to secure the agency of individuals without reducing ethical obligation to the level of legal obligation. For example, as has been noted, and as we will further discuss when we turn to O'Neill's alternative Kantian constructivism, reciprocity (while not irrelevant) is not fundamental to Kantian morality or justice.

38 Rawls stressed that it must be stable "for the right reasons", i.e. through the free endorsement of the principles by all citizens.
strategy in the idea that a consensus could form around a minimal and (hopefully) uncontroversial account of the specifically political values and ideals that underpinned a liberal democratic society.

In what follows I will first examine the “freestanding” conceptions that replaced Rawls’s earlier philosophical conceptions, in particular of the person, freedom, equality and autonomy, before turning to his revised approach to justification, legitimacy and stability. Central to all this is his move away from supposedly contested philosophical or “metaphysical” claims towards what he called “purely political” minimal claims that were concerned only with answering the most basic questions of justice that arise within a bounded liberal democratic society.

1.2.2. Freestanding conceptions: Freedom and equality

A Theory of justice posited a conception of the person as a moral person characterised as free, equal and rational. This freedom and equality was linked to the fundamental nature of human beings, and consequently it was suggested that the realisation of this freedom and equality through the principles of justice as fairness was also a fundamental aspect of humankind’s “good”. As we have seen, Rawls came to believe this “Kantian” element was too much at risk of accusations of metaphysical suppositions. In revising his account of justice he sought to rearticulate the account in such a way that there could be no such accusations. Yet in doing this he was also concerned that the guiding conceptions of freedom and equality not be lost, as these were fundamental. Rawls therefore outlined a conception of the person as citizen, which he hoped could be accepted by all citizens of a liberal democratic polity.

In Political Liberalism Rawls offers three ways for viewing citizens (not persons) as free. Firstly, they have a conception of the good, in that they have a private identity separate from their public life and may freely pursue the values and ideals expressed by that identity (subject to reasonable limitations). This identity is not equivalent to their public identity, and hence they are also free in the sense that
any changes to this private identity do not effect their status and standing in relation to other citizens, i.e. their public identity (cf. PL, 30).  

Secondly, citizens are to be seen as what Rawls calls “self-originating sources of claims” (PL, 32). This is best understood in contrast to less liberal societies, where persons are only recognised as entitled to make claims on the basis of certain ascribed duties and obligations derived from their status or role in that society (PL, 33). So citizens under Rawls’s account are free in the sense that their claims may be derived from their own particular conception of the good. The third way in which citizens are free, then, is in respect of their capacity to revise their conceptions of the good. This aspect of their freedom implies a limitation on their conceptions of the good; it is assumed that citizens are free to revise their ends and therefore they acquire a social obligation to do so if the pursuit of these ends demands more than the resources available to them allows (provided resources are distributed according to a just basic structure (cf. PL, 33-4)).

As for citizens viewing themselves as equal, this is achieved on the basis that each is presumed to have the capacity to both formulate and revise a conception of the good, i.e. to be both rational and reasonable. While citizens will have the requisite capacities to greater or lesser degrees, each is presumed to meet a minimum threshold that allows them to participate fully in the social life of a liberal democratic society (PL, 19). On this basis they may view each other as equal.  

Given Rawls’s distancing from metaphysical claims about the nature of persons, an account is needed of the foundation of this conception of the citizen as free and equal. What is of most importance for this account is that it is the citizen, and not  

---

40 Rawls considers cases in which persons lack the ability to be what he calls “normal and fully cooperating members of society” to be special cases which require supplemental consideration. He does not include them because he seeks a simplified framework that offers a clearer picture of how “normal” social interactions should be regulated through a basic social framework. Rawls, “Justice as Fairness: Political not Metaphysical”, p. 234. This position has of course attracted controversy, and there are questions to be raised regarding what should constitute a “normal” member of society. I cannot address these issues here, however, when we examine O’Neill’s alternative Kantian account of justice, we will see that her account potentially offers a more satisfactory way of accommodating a spectrum of human capacities for reason and action.
the person, to whom these attributes of freedom and equality are ascribed. Some comments are instructive. In his 1985 article, "Justice as Fairness: Political not Metaphysical", Rawls states that "[j]ustice as fairness starts from the idea that society is to be conceived as a fair system of cooperation and so it adopts a conception of the person to go with this idea" (emphasis mine). Given that society is seen as fair and cooperative, it requires a conception of the citizen that reflects this understanding of society. Freedom and equality are then necessary because they give expression to the citizen's capacity to take part in society in the two ways relevant to fairness and cooperation. Fairness is reflected in each citizen's conceptions of the good being given equal weight, and cooperation is reflected in the capacity to revise that concept in the interests of fairness. In this way Rawls defends his conception of the citizen as purely political; it only concerns how these capacities affect and are affected by the basic social structure, i.e. what sort of social obligations, rights and liberties derive from these considerations.

Rawls claims this conception of the person adopted by justice as fairness does not originate from any philosophical doctrine, rather it originates in the public political culture of a democratic society. In "Justice as Fairness: Political not Metaphysical" he states that "the conception of person as having the two moral powers, and therefore as free and equal, is also a basic intuitive idea assumed to be implicit in the public culture of a democratic society". Further on he adds, "When we describe a way in which citizens regard themselves as free, we are describing how citizens actually think of themselves in a democratic society should questions of justice arise. In our conception of a constitutional regime, this is an aspect of how citizens regard themselves". Justification for such a conception then, while apparently linked to the role it plays in justice as fairness, derives from it being the intuitive and implicit conception that persons in a democratic society already and actually appeal to when debating matters of basic justice. Only citizens can assert their freedom and equality, on the basis of their capacity and entitlement to participate in a liberal democratic society. Justification is only available to citizens; it is questionable whether persons abstracted from the particularities of a liberal

democratic regime can draw on these implicit ideals.

Given that freedom and equality have been detached from understandings of the self and reconfigured as features of citizens that are relevant to the basic structure of a society based on fairness and cooperation (specifically liberal democratic societies), Rawls therefore also needed to re-articulate his understanding of autonomy: he now needed to ground the claim to self-legislation in something other (and less divisive) than the free and equal self expressing its true nature through the principles of justice.

**1.2.2.1. Rational autonomy: freedom to pursue ends**

According to Rawls, there are two ways in which the parties to the original position, and correspondingly the citizens of a democratic polity, are rationally autonomous. Citizens are free to pursue their conceptions of the good, and, correspondingly, the parties are free to choose whatever principles best achieve this aim for those they represent. Citizens also are motivated “to secure their higher-order interests associated with their moral powers”, and correspondingly, parties take these interests into account (PL, 72-3).

This describes citizens' rational autonomy as follows; firstly, because citizens pursue their own conception of the good, the selection of principles is not thought to be guided by any heteronomous source of authority. That is, accordingly to Rawls, the parties are not “bound by any antecedently given principles of right and justice”, or, “[p]ut another way, they recognize no standpoint external to their own point of view” (PL, 73). They are “self-legislating” in the sense that it is their self and their conception of the good that guide the selection of principles.

Secondly, because citizens are motivated by a higher-order interest (to secure the primary goods needed to pursue their conception of the good), they are not

---

44 Citizens are “free” in an explicitly non-Kantian sense: they have negative freedom to exercise their preference’s in their non-public life.
motivated by a heteronomous source of authority. Rather the source, their "highest-order" interest, is internal to their self. As in *A Theory of Justice*, Rawls is careful to distinguish higher-order interests from general desires, which he sees as heteronomous sources of authority, and here we may trace the shift from, and revision of, the notion of autonomy first articulated in *Theory*. In his first book Rawls, argued that higher-order interests are distinguished from desires because they relate to the expression of the fundamental nature of the person. The principles of justice expressed "our" fundamental nature as free and equal moral persons; they derived from the self's interest in expressing itself, and moral persons acting from these principles were therefore autonomous.

In *Political Liberalism* the concern to express freedom and equality is retained, but it is re-articulated; claims about the fundamental nature of persons are dropped and instead what is emphasised is the connection between freedom and equality and the moral powers and capacities of citizens. Freedom and equality are now grounded in the capacities that enable citizens to pursue a conception of the good, and the primary goods needed for it, in accordance with principles of justice; citizens now value freedom and equality because they value capacities to pursue conceptions of the good in accordance with principles of justice (PL, 77). Where, in *A Theory of justice*, expressing our fundamental nature was seen as a "good", that value has been replaced by the value of pursuing a conception of the good, and freedom and equality are valuable insofar as they relate to that capacity. This re-articulation is expressed by Rawls as follows: "[t]he parties are trying to guarantee the political and social conditions for citizens to pursue their good and to exercise the moral powers that characterize them as free and equal" (PL, 76). Citizens are rationally autonomous because they act out of a higher-order interest connected to the rational pursuit of their own ends, and freedom is a function of this pursuit.
1.2.2.3. Full autonomy: equality as a constraint on the freedom to pursue ends

Citizens, aside from having a conception of the good, also view themselves as equal because they recognise that others have the powers and capacities to be participants in social life. This takes us to the level of “full autonomy”, conceived as political and not metaphysical (PL, 77). We noted that freedom and equality are now grounded in the capacities that enable citizens to pursue a conception of the good in accordance with principles of justice. Freedom relates to the capacity to pursue a conception of the good, and equality relates to how this pursuit is regulated by the principles of justice. Where “rational autonomy” allows citizens, as free citizens, to deliberate about what principles are in their own rational interest, “full autonomy” constrains this deliberation so that the principles reflect both citizens’ interests and also how those interests are affected by others’ equally relevant claims. Rational autonomy only becomes full autonomy when a conception of equality is operationalised. Since equality reflects how citizens perceive their higher-order interest in pursuing a conception of the good, it is autonomous, and not heteronomous. Full autonomy combines the (politically interpreted) freedom, equality, rationality and reasonableness of citizen’s and expresses these political conceptions via acceptance and endorsement of a political conception of justice.

1.2.3. Freestanding conceptions: a political conception of justice

Rawls’s guiding concern was to justify a conception of justice for a plurality of diverse persons holding many comprehensive doctrines. This meant that his account of justice could not rely on any one comprehensive doctrine as its foundation; it could not be understood as the political application of a much wider doctrine that encompassed values and ideals that applied to other domains of human life. Rather it would have to be detached from all such comprehensive conceptions.
A further requirement was that such a conception must not conflict with any reasonable comprehensive doctrine; it must function as a "module" capable of being embedded within a wider doctrine. While this requirement places demands on comprehensive doctrines themselves (by asserting that reasonable doctrines were those that would endorse and embed within them such a conception) it also has implications for the political conception itself. Specifically it implies that such a conception must be limited in particular ways so as not to encroach into domains where its ideals and values might come into conflict with a plurality of diverse value systems.

This requirement implied that such a conception must be limited to the political domain, to resolving questions that concerned the institutional framework within which citizens carried out their interactions with one another. Rawls called this framework the "basic structure". He suggested that because this framework had a deep and pervasive influence over citizens' lives, which began at birth and lasted across their lifetime (bracketing issues of migration), it was crucial that such a framework be worked out in a way that allowed for all reasonable conceptions of the good to be pursued. As such, no one conception could dictate the design of the basic structure. Therefore, according to Rawls, only a political conception – that eschewed other ethical and moral questions – could justify a framework of basic justice. Because a political conception concerned itself only with questions of how a diversity of free and equal persons could pursue a variety of conceptions of the good within one society, in his view it alone was capable of answering such questions in relation to the basic structure. The scope and limits of a political conception were crucial in both directions; to attempt to justify more would risk conflict with wider doctrines, and to attempt less would mean an unacceptable encroachment of comprehensive doctrines into that area of the political sphere that dealt with basic justice, rights and liberties.

This brings us to the question of content. Political Liberalism describes the content of a political conception of justice as those "ideas implicit in the public political culture" (PL, 100) of a liberal democratic society, or more specifically: "[s]ociety's main institutions, and their accepted forms of interpretation, are seen as a fund of
implicitly shared ideas and principles" (PL, 14). In his earlier 1989 article, “The Domain of the Political and Overlapping Consensus”, Rawls offered more insight into the substance of such content. He states that “it has the kind of content we historically associate with liberalism”, and, further, “[t]he conception consists in a view of politics and the kinds of political institutions which would be most just and appropriate” (taking into account what Rawls considered to be the facts that constituted the modern context of justice, e.g. reasonable pluralism owing to the burdens of judgment). He also states that a particular set of basic political and civil rights and liberties, and their order of priority, forms the content of a political conception of justice. He describes justice as fairness as one particular “formulation of highly significant (moral) values that properly apply to basic political institutions; it gives a specification of those values which takes account of certain special features of the political relationship.” For example, justice as fairness describes the political relationship as a fair system of social cooperation, expressed through the criterion of reciprocity, and so articulates a specific set of rights and liberties to reflect this description. While clearly Rawls continued to defend justice as fairness as the most just political conception, he also insisted that there were other specific articulations of such political ideals around which a just consensus could (or would) develop.

The limiting of the content of a political conception of justice to such basic rights and liberties was essential for carrying out the limited purpose of a political conception. In “The Domain of the Political”, Rawls noted that the idea of a political conception lead directly to “the idea of ‘constitutional essentials’, and that a ‘specification of the basic rights and liberties of citizens... falls under those essentials’.” The purpose of a political conception of justice was then to provide “a reasonable framework of principles and values for resolving questions concerning these essentials”. We will return to the relationship between the

46 John Rawls, “The Domain of the Political”, p. 166.
48 John Rawls, “The Domain of the Political”, p. 166.
49 John Rawls, “The Domain of the Political”, p. 166.
content of a political conception and constitutional essentials when we examine his account of public reason, as it is the values and ideals of a political conception that form the content of public reason.

1.2.4. Justification and the political conception

Rawls offers three stages of the justification of the political conception of justice, which can be understood to entail, respectively, the principles and ideas of practical reason (1), reflective equilibrium (2) and finally the overlapping consensus (3). These then correspond to the considered convictions of moral persons (as we have already seen) (1), the considered convictions of citizens (2) and finally to the reasonableness of citizens in the public sphere of a liberal democratic polity (3). In *A Theory of Justice*, the considered convictions upon which the second method of justification (besides the contract foundation) rests are described in the abstract as those of moral persons, however in *Political Liberalism* the moral person is replaced by the citizen of a liberal democratic polity. More specifically, Rawls emphasises that it is the conception of the person as citizen, which is *and has always been* the appropriate conception.

By examining the stages of justification I hope to draw further attention to the constitutive assumptions Rawls makes with regard to the relationship between the principles and ideas of practical reason and substantive principles of justice, and the relationship he establishes between justification, stability and legitimacy, specifically why and to what extent justification rests on an "overlapping consensus".

1.2.4.1. Considered convictions in Rawls’s account of practical reason

The first stage of justification presented by Rawls is what he termed "pro tanto justification"; the political conception is justified insofar as the arguments presented, those arguments relating to the principles and ideas of practical reason,
appear on the face of it to justify the theory. Practical reason as interpreted by Rawls entails both rational and reasonable principles; reasoning about ends and reasoning about the coordination of ends (where a plurality of such exist). Of those reasonable principles it is concern for impartiality of judgment that is fundamental. As we have seen, unlike Kant, Rawls rejects the idea that "practical reason itself can constitute the moral or political values" (PL, 99). Rather, in what Rawls takes as opposition to Kant, the principles of practical reason are those considerations that persons intuit as relevant to justice, that the appropriate point of view for the deliberation and selection of principles of justice is a suitably impartial perspective. The ideas of practical reason then are those that correspond to and complement the principles of practical reason: the idea of the person as free and equal citizen possessing the two moral powers, and of society as a well-ordered, fair system of social cooperation (PL, 107). According to Rawls, such ideas follow necessarily from reflection on the principles of practical reason:

[T]he principles of practical reason are expressed in the thought and judgment of reasonable and rational persons and applied by them in their social and political practice... they characterize the agents who reason and they specify the context for the problems and questions to which principles of practical reason apply. (PL, 107. Emphasis mine)

It is in this sense that pro tanto justification is capable of reflecting the considered convictions of moral persons now understood as citizens. The principles (and hence ideas) of practical reason are held by Rawls to be accessible to all moral persons. However, what is significant now is that Rawls emphasises how necessity and complementarity define the relationship between the principles and the ideas of practical reason, or more specifically, that notions of citizenship and liberal democratic society are un-detachable from the principles of practical reason. Rawls moves from the considered convictions of moral persons, in the abstract, to specifically the practical reasoning of citizens by claiming that this conception of the person (and society) is wedded to the principles of practical reason. Practical

50 For further reading on the connection between justification and the principles of practical reason in Rawls's theory, see Rainer Forst, Contexts of Justice, pp. 154-200. Forst defends the internal justification of Rawls's principles through their derivation in the principles of practical reason, while at the same time critiquing what he sees as Rawls's conclusions about the substantive content of principles of justice.
reason without a specified agent possessing the reasonable and rational powers of practical reason, and without a notion of society that reflects and extends this conception, "would have no point, use or application" (PL, 108).

Unlike Kant, Rawls does not see practical reason itself as providing its own premises for reasoning about justice, rather it is the practical reason of embodied agents, as citizens of a liberal democratic polity, that provides the foundation, the "fixed ideas", for a constructivist account of justice. It is not pure practical reason; rather it is the reason of partially contextualised agents. Reason must always be "from somewhere" because this context specifies the practical issues for which practical reason is deployed (PL, 116). It is then political and not metaphysical: it is emphasised now that the ideas of practical reason are those implicit in the public culture of modern liberal democratic societies. It speaks to those who already accept (or are ready to accept) that the principles of practical reason automatically imply conceptions of person and society defined respectively as citizen and liberal democratic regime.

As we saw earlier, the only vindication apparently offered for this derivation of substantive conceptions of person and society from the principles of practical reason is that practical reasoning represents the form of "common human reason" that Rawls identifies with Kantian practical reason (though in a political and not "pure" sense):

In Kant's doctrine, it is the point of view of such persons as members of a realm of ends. This shared point of view is possible since it is given by the categorical imperative which represents the principles and criteria implicit in their common human reason. Similarly, in justice as fairness it is the point of view of free and equal citizens as properly represented. (PL, 115)

As mentioned above, Rawls claims that "[t]here is no such thing as the point of view of practical reason as such" (PL, 116). Rather, because his account is political, the principles of practical reason represent the point of view of citizens. It is therefore necessary that the basis of objectivity (and hence of public justification) must be the shared principles and ideas of practical reason implicit in the public political culture. This claim in turn relates back to Rawls's concerns over
reasonable pluralism; his fears that reaching agreement between a multiplicity of
doctrines was not possible by reasoned discourse alone, and his claim that only a
substantive *shared basis* for reasonable agreement could provide the foundations
for a stable political consensus. While, as we shall see, the principles of a liberal
political conception of justice form the content of a public reason by which citizens
could reason together, this stage is reached by appeal to a more fundamental
shared basis grounded in the principles of practical reason:

Of course, given the many obstacles to agreement in political judgment even among reasonable
persons, we will not reach agreement all the time, or perhaps even much of the time. But we should
be able at least to narrow our differences and to come closer to agreement, and *this in the light of
what we view as shared principles and criteria of practical reasoning*. (PL, 118. Emphasis mine)

The form of objectivity appropriate to political constructivism is grounded in
shared principles of practical reason. In turn, public justification is grounded in the
objective point of view. And given that, as we saw above, the point of view of free
and equal citizens is the appropriate understanding of objectivity, we see that the
principles of practical reason reflect this point of view. Common human reason is
now common citizens’ reason. Also it is apparent now that the overlapping
consensus, which forms the shared basis of public justification, is *derived from the
principles of practical reason*, understood as citizens’ reason.

Rawls makes several comments regarding this point in his discussion of practical
reason and constructivism. He writes:

>[T]he constructivist’s conjecture is that *the correct model of practical reason as a whole will give
the correct principles of justice on due reflection*... Thus, if the procedure can be correctly
formulated, *citizens should be able to accept its principles and conceptions along with their
reasonable comprehensive doctrines*. The political conception of justice can then serve as the
focus of an overlapping consensus. (PL, 96-7. Emphasis mine)

The basic claim Rawls makes is that if citizens agree on the principles and ideas of
practical reason, and if the procedure properly reflects these principles and ideas,
then it follows almost directly that citizens should be able to agree on the
principles of justice constructed via this procedure. Agreement or consensus around a set of liberal principles is possible because the overlapping consensus simply expresses what citizens already share with each other in their practical reason; they are likely to come to agreement because the content of the consensus merely takes those principles and ideas they already share and gives them substance and form through the constructivist procedure.

There appears an apparent dichotomy between Rawls's claim of the universal-moral validity of liberal principles and their endorsement by various "reasonable" traditions from a plurality of different backgrounds. How is the claim to a "shared point of view" reconcilable with the apparent "overlap" of a diversity of points of view?

According to Rainer Forst, the problem is that Rawls does not emphasise enough the moral-universal justification of the theory; he must not subordinate the "moral requirement" to the "practical requirement" of being compatible with ethical conceptions of the good (CJ, 183). He suggests that Rawls's appeal to the support of ethical doctrines risks undermining the moral justification of his theory:

Contrary to the view that the theory can do without a strong claim to moral validity, it will become evident that the theory can be neither ethical nor political "in the wrong way," as Rawls claims, only if it can fall back upon a conception of the moral person that can be justified solely as an "idea of practical reason." The conception of justice can defend its moral priority over ethical conceptions only if it is more than a rational compromise and less than an ethical doctrine, that is to say, if it is a morally justified conception. (CJ, 183-4)

The situating of moral justification between "a rational compromise" and "an ethical doctrine" flags some important considerations regarding the relationship between moral-universal justification and ethical-contextual validity in Rawls's theory. What does this dichotomy imply for the status of ethical doctrines in Rawls's account? If they can agree on a set of liberal values based on shared practical reason, what remains that is unique and distinct in the normative core of each tradition? If their commitment to principles of justice derives from a shared practical reason that is embedded as a module within all "reasonable" traditions, it
seems that this claim must either undermine the validity claims of ethical doctrines or else reduce their normative core to a modus vivendi. If the principles of justice are endorsed from different sources (e.g. Christianity deriving principles of impartiality from the Golden Rule) then it seems just luck that they settle on the same principles. However Rawls’s argument goes further and suggests that all reasonable persons have a shared capacity for moral reasoning.

But it seems that the two are not compatible: either we share a capacity for moral reasoning, and on that basis we should be capable of engaging in an exchange of ethical reasons, or we do not share such a capacity and ethical doctrines have little or nothing to say to each other beyond a modus vivendi. Yet Rawls rejects the former and denies that his theory endorses the latter; he rejects the assertion that his account is a modus vivendi yet also rejects the possibility of a fruitful exchange of reasons between a plurality of worldviews. Again it seems that he presents two opposing views of moral reasoning: on the one hand a universally shared capacity that opens up possibilities for living together in just societies, and on the other distinct forms of reasoning from mutually impenetrable ethical traditions that can only co-exist by “avoiding” engagement with each other.51

1.2.4.2. Reflective equilibrium and citizens’ considered convictions

The second stage of justification, “full” justification, moves from internally coherent justification to justification through external reflective endorsement. As in the process of reflective equilibrium described in A Theory of Justice, the person (now emphasised as citizen and not moral person), reflects on whether their own convictions (now fleshed out in comprehensive doctrines) are in line with, or can

---

51 Rawls describes his approach as follows: “As I have said, we must apply the principle of toleration to philosophy itself. The hope is that, by this method of avoidance, as we might call it, existing differences between contending political views can at least be moderated, even if not entirely removed, so that social cooperation on the basis of mutual respect can be maintained. Or if this is expecting too much, this method may enable us to conceive how, given a desire for free and uncoerced agreement, a public understanding could arise consistent with the historical conditions and constraints of our social world.” See Rawls, “Justice as Fairness: Political not Metaphysical”, p. 231.
be brought into line with, the principles of justice constructed through a modelling of the shared principles and ideas of practical reason.

Rawls describes this as a process whereby “the citizen accepts a political conception and fills out its justification by embedding it in some way into the citizen’s comprehensive doctrine as either true or reasonable, depending on what that doctrine allows” (PL, 386). Though the political conception is freestanding, at this stage it is seen whether it is also capable of being inserted into citizens’ wider ethical frameworks. This stage is described as “full” justification because, Rawls says, “[s]ome may consider the political conception fully justified even though it is not accepted by other people. Whether our view is endorsed by them is not given sufficient weight to suspend full justification in our own eyes” (PL, 386. Emphasis mine). Justification at this level is complete only from the perspective of the individual citizen who has not yet brought their convictions forward into the public sphere.

To understand what precisely is happening at this level, it is helpful again to recall what precisely is being justified; it is a set of liberal principles constructed via a hypothetical thought experiment contrived to determine what suitably described agents would choose for themselves as principles of justice in a process of rational deliberation constrained only by considerations relevant to justice. The hypothetical scenario sets out principles that would be endorsed by all under reasonable conditions. At the level of full justification then, what is given consideration is whether each individual citizen endorses principles that have already been shown, through internal justification, to be those that would be endorsed by all. This stage of justification attempts to align the hypothetical intersubjective acceptance of principles of justice with the actual acceptance, at the individual citizen’s level, of principles that would be intersubjectively endorsed in a hypothetical scenario. What is not shown or justified here is whether such principles receive actual intersubjective endorsement or acceptance from the whole public of a specified liberal polity made up of “reasonable” citizens.
1.2.4.3. Overlapping consensus and citizens’ reasonableness

At the third stage of justification, that of “public justification”, it is now all citizens collectively that carry out the process of justifying the political conception of justice by embedding its principles within their own comprehensive doctrines. Rawls describes this as a process whereby “reasonable citizens take one another into account as having reasonable comprehensive doctrines that endorse that political conception”, a process of “mutual accounting” (PL, 387). However, citizens do not take the content of each other’s comprehensive doctrines into account, only the fact of their existence; it is the fact of reasonable pluralism itself which plays the key role (cf. PL, 387). Justification does not end, as in A Theory of Justice, with “full” justification via the individual’s considered convictions being held in reflective equilibrium with the principles of justice, it is extended now to a collective, “general and wide reflective equilibrium” which, though not final, is “the best justification... we can have” (PL, 388).

This move from individual to collective reflective equilibrium reflects the move from citizens’ considered convictions to citizen’s reasonableness. Now reflective equilibrium it is not merely reflecting on (from the perspective of a citizen) whether a certain set of liberal principles coheres with one’s own convictions about justice, it is also recognising that there exist many reasonable considered convictions and that the principles one endorses must also be acceptable to those other reasonable positions. It is not merely an aggregation of subjective reflective equilibria; it is now also, to at least some extent, an intersubjective process. This stage moves past “full” justification and hopes to align the hypothetical intersubjective acceptance of principles of justice with the actual intersubjective endorsement of a set of liberal principles by the whole public of a specified liberal polity made up of “reasonable” citizens. The overlapping consensus is the outcome of a successful undertaking of such a process.

---

52 While I describe this as intersubjective, it must be noted that citizens do not look inside the content of other comprehensive doctrines. Therefore this is vastly distinct from a properly discursive or deliberative understanding of intersubjectivity, such as that of the discourse ethicist Jürgen Habermas.
Rawls views this last stage of justification as indispensable to the justification of a political conception of justice; it is only when a public basis of justification has been established (the overlapping consensus), that questions of stability and legitimacy can be addressed (cf. PL, 388). The issue of stability is not an additional consideration; rather it goes to the heart of political liberalism. In the early pages of Political Liberalism Rawls asserts that “[t]he problem of political liberalism is to work out a conception of political justice for a constitutional democratic regime that the plurality of reasonable doctrines might endorse” (PL, xviii. Emphasis mine). On the following page he again stresses that “the aim of political liberalism is to uncover the conditions for the possibility of a reasonable public basis of justification” (PL, xix. Emphasis mine). Further on he continues that “to serve as a public basis of justification for a constitutional regime a political conception of justice must be one that can be endorsed by widely different and opposing though reasonable comprehensive doctrines” (PL, 35. Emphasis mine).

An obvious question that presents itself at this stage, and one that has been a source of controversy, is the appropriateness of linking what appears to be a key stage of justification to the question of stability; why should we, or perhaps more pressingly, why does Rawls, consider the stability of a regime to determine the justification of principles of justice? This question is of utmost importance as it is this question of stability that necessitates the overlapping consensus, which, as we will see, forms the content of Rawlsian public reason.

Rawls connects public justification to both stability and legitimacy because, respectively, stability shows that the principles are actually endorsed and internalised by the majority of a liberal polity, and the principle of legitimacy guarantees that principles, policies and laws are justified to all citizens of a particular polity. It seems, on the face of it, that questions of stability and legitimacy relate in rather different ways to justification. Questions of legitimacy appear more appropriately related to justification as they relate to what can be accepted. In contrast, stability relates to what is actually accepted. Yet Rawls takes both to relate to the public basis of justification. Some comments from Rawls help
elucidate why he takes stability and legitimacy to apply equally to the same level of justification. Of stability (via the overlapping consensus), he writes:

Political liberalism makes no attempt to prove, or to show, that such a consensus would eventually form around a reasonable political conception of justice. The most it does is to present a freestanding liberal political conception that does not oppose comprehensive doctrines on their own ground and does not preclude the possibility of an overlapping consensus for the right reasons. (PL, xlv-xlvi. Emphasis mine)

And on the same point, regarding the possibility of consensus or agreement, Rawls noted, in "Justice as Fairness: Political Not Metaphysical", that the "kernel" of such a consensus was the "shared intuitive ideas" of citizens in a liberal polity. Of legitimacy, he writes:

[O]n matters of constitutional essentials and basic justice, the basic structure and its basic policies are to be justifiable to all citizens, as the principle of political legitimacy requires. We add to this that in making these justifications we appeal only to presently accepted general beliefs and forms of reasoning found in common sense, and the methods and conclusions of science when these are not controversial. (PL, 224. Emphasis mine)

The desire for stability implies a concern for the possibility of consensus, with such possibility being grounded in a shared basis for further agreement. Similarly with the question of legitimacy, justification is grounded in a shared basis of agreement. On this interpretation then, questions of what can be accepted and what would or is actually accepted are all grounded in this shared basis of agreement. For our purposes, to show how and why Rawls links stability to justification and legitimacy, what matters here is that showing that something can be accepted must follow the same justificatory path as showing that it possible that something will be accepted. Rawls evidently takes the two to be interchangeable, or, at least, he does not distinguish between something that it is possible to accept, and the possibility that something will be accepted.54

54 Habermas has criticised Rawls on this point: "In my view, Rawls must make a sharper distinction between acceptability and acceptance. A purely instrumental understanding of the theory is already invalidated by the fact that the citizens must first be convinced by the proposed conception of justice before such a consensus can come about. The conception of justice must not be political in
We may examine this point further by considering what Rawls means when he speaks of the "possibility" of something being accepted, given that we now know this meaning relates to both stability and legitimacy. Rawls makes two assumptions with regard to what can possibly be accepted. First, that no claims derived from particular comprehensive doctrines could possibly be accepted by a plurality holding diverse worldviews, and, second, that only principles adopted from an appropriately described impartial perspective (the objective perspective of citizen's shared practical reason) could possibly be accepted. I would argue however that the appropriate interpretation of these assumptions is not that principles that do not meet these conditions could not possibly be accepted in the actual, modal sense of possibility, but in the sense that it is not probable that they will be accepted. The underlying assumption is that citizens will not be motivated to accept principles that do not meet these conditions. Rather than interpreting possibility in terms of capacity and agency, Rawls is attempting to show the possibility of motivating persons to adopt just principles. If we enquire as to why he interprets possibility in these terms, a potential and plausible response is that it is because his theory of justice begins with a preference-based approach to agency that takes hypothetical consent to just principles to be the aim of a constructivist account of justice. Since Rawls begins with an account of the principles persons in an ideal situation would accept, he must then consider at a later stage in his theory why actual persons would also accept those same principles. It is for this reason that his three stages of justification move from hypothetical intersubjective endorsement, to actual subjective endorsement, and finally to actual intersubjective endorsement (or citizen's reasonableness).

---

the wrong sense and should not merely lead to a modus vivendi. The theory itself must furnish the premises 'that we and others recognize as true, or as reasonable for the purpose of reaching a working agreement on the fundamentals of political justice.' But if Rawls rules out a functionalist interpretation of justice as fairness, he must allow some epistemic relation between the validity of his theory and the prospect of its neutrality toward competing worldviews being confirmed in public discourses." Habermas, "Reconciliation Through the Public use of Reason", p. 122.

55 “Modal” is term used by O'Neill and which is crucial both to her critique of Political Liberalism and Rawls's account of public reason and to understanding her particular interpretation of Kant's ethics. See O'Neill, "Political Liberalism and Public Reason", in particular, pp. 416-427.
What has become evident from examining the stages of justification of the political conception of justice is that, firstly, the content of the overlapping consensus is necessarily a set of substantive and particular liberal principles that Rawls believes derive directly from the principles and ideas of practical reason. Secondly, the actual consensus on this set of principles is crucial to justification as a result of the starting points Rawls sets out in his particular constructivist approach.

Recall that the primary concern of Political Liberalism (aside from the justice of principles) is stability and hence the possibility (understood as probability) of agreement on principles of justice within liberal constitutional democracies. Of course, Rawls’s concern for agreement extends not only to the initial construction of just principles but also to their ongoing role in public political debate. Beyond the public justification of principles of justice there remain important political questions that require some method for attempting to achieve agreement or consensus. For this reason the next task of the political conception of justice, according to Rawls, is to work out an application of the overlapping consensus to the domain of political debate on basic political questions.

1.2.5. Public Reason: the public basis of justification in liberal democracies

Questions of stability, and hence possibilities of consensus on just principles and agreement on political matters, guide Rawls’s account of public reason. Beyond the establishment of principles of justice, a liberal democracy must also contain a public sphere where matters of first importance are discussed, debated and decided upon. It was Rawls’s conviction that if public debate was allowed to proceed in the manner of wholly free and open discourse, then the inevitable outcome would be what he termed “irreconcilable conflict”:

[A] basic feature of democracy is the fact of reasonable pluralism, the fact that a plurality of conflicting reasonable doctrines, religious, philosophical, and moral, is the normal result of its culture of free institutions. Citizens realize that they cannot reach agreement or even approach mutual understanding on the basis of their irreconcilable comprehensive doctrines. In view of this, they need to consider what kinds of reasons they may reasonably give one another when
fundamental political questions are at stake. I propose that in public reason comprehensive doctrines of truth or right be replaced by an idea of the politically reasonable addressed to citizens as citizens. (PL, 441. Emphasis mine)

Rawls was unequivocal on the possibilities for agreement where debate proceeded with the inclusion of those convictions of persons that belonged to what he designated the private sphere. The burdens of judgment meant that even conscientious and reasoned debate could never guarantee a consensus on non-political issues (cf. PL, 54-5). Reasonable persons were, therefore, obligated to respect and tolerate other doctrines as reasonable (though not true) so long as those other doctrines were willing to do likewise. Since free debate between conflicting worldviews would engender irresolvable disagreement, public debate in a democratic society must seek another basis for reasoned political agreement. As we have seen, it was Rawls’s belief that the appropriate basis for political agreement was a public basis of justification grounded in the shared reasoning of citizens; public justification, according to Rawls, “proceeds correctly from premises we accept and think others could reasonably accept to conclusions we think they could also reasonably accept” (LP, 155).

Public justification must be limited by the criterion of reciprocity, and, correspondingly, by the principle of liberal legitimacy. Because political power

---

56 By “could reasonably accept”, Rawls meant that the reasons offered stemmed from the shared principles and ideas of practical reasons, and not “could accept” in the modal sense.

57 On reciprocity and legitimacy (legitimacy as the offering of reasons that could be accepted by others): “Citizens are reasonable when, viewing one another as free and equal in a system of social cooperation over generations, they are prepared to offer one another fair terms of social cooperation (defined by principles and ideals) and they agree to act on those terms, even at the cost of their own interests in particular situations, provided that others also accept those terms. For these terms to be fair terms, citizens offering them must reasonably think that those citizens to whom such terms are offered might also reasonably accept them. Note that “reasonably” occurs at both ends in this formulation: in offering fair terms we must reasonably think that citizens offered them might also reasonably accept them. And they must be able to do this as free and equal, and not as dominated or manipulated, or under the pressure of an inferior political or social position. I refer to this as the criterion of reciprocity (PL, xlii). And on reciprocity and public justification: “When political liberalism speaks of a reasonable overlapping consensus of comprehensive doctrines, it means that all of these doctrines... support a political conception of justice underwriting a constitutional democratic society whose principles, ideals and standards satisfy the criterion of reciprocity... On the other hand, comprehensive doctrines that cannot support such a democratic society are not reasonable. Their principles and ideals do not satisfy the criterion of reciprocity and in various ways they fail to establish the basic liberties... Since the criterion of reciprocity is an essential ingredient specifying public reason and its content, political liberalism rejects as unreasonable all such doctrines.” (PL, 482-3)
implied the coercive power to enforce laws and policies upon fellow citizens, the use of political power must meet the criteria of reciprocity and legitimacy. These demanded that each citizen was obligated to offer justification for their reasoning that all other citizens could in principle accept:

(S)ince political power is the coercive power of free and equal citizens as a corporate body, this power should be exercised only when constitutional essentials and basic questions of justice are at stake, only in ways that all citizens can reasonably be expected to endorse in the light of their common human reason. (PL, 139-40)

Or: citizens “should be ready to explain the basis of their actions to one another in terms each could reasonably expect that others might endorse as consistent with their freedom and equality” (PL, 218). As we have already seen, because Rawls begins with hypothetical consent and therefore links possible acceptance with probable acceptance, it is implied that reasoning begins from shared premises and therefore proceeds to acceptable conclusions. As such reasoning must begin from shared premises, the contents of comprehensive doctrines were inadmissible as potential public reasons. Rather it is the overlapping consensus on liberal principles, understood as the shared principles and ideas of practical reason, given form and content through the constructivist procedure, that must serve as the basis for public reason.

Rawls asserts that the role of the citizen in the public political forum is defined as an ideal; the “duty of civility” is defined as an ideal of democracy (PL, 252). It is the duty of citizens, in their capacity as bearers of democratic powers, to exercise such powers always with regard to the limits of public reason; citizens are obliged never to introduce into debate concerning matters of basic justice or constitutional essentials any non-political, and hence (on Rawls's interpretation) unshareable, reasons or values. Much of the reference Rawls makes to this ideal and its relation to public reason, especially to citizens' “duty” and “conduct”, imply that there is a great emphasis placed on the role of the citizen in distinguishing for themselves, before they enter the public sphere, their public and private identities (PL, 252-3). This reading is emphasised by Rawls's distinction between an “inclusive” and “exclusive” view of public reason. While under an exclusive view, citizens are
required to never introduce any aspect of their comprehensive doctrines into public debate, on an inclusive view they may, as in the case of deep religious division, explain publically in what way their comprehensive doctrines affirm the political principles (cf. PL, 448-9). The burden on the citizen is emphasised here because we see that, even at its most flexible, public reason merely allows for citizens to explain in what way the political values are embedded in their comprehensive doctrines, and not to present any other values that are not already given by the political conception. At its most flexible, public reason still demands a considerable separation of identities on the part of the individual citizen. Furthermore, this flexible view is only offered as a concession to nonideal circumstances, it is certainly a less than normative view for Rawls (cf. PL, 252).

The role of the duty of civility is also decisive in terms of the entire account of the stability of well-ordered institutions; “the political conception and its ideal of public reason are mutually sustaining, and in this sense stable” (PL, 252).

As for the content of public reason, we saw, when examining the political conception as freestanding, that the content of public reason was drawn from “a fund of implicitly shared ideas and principles” (PL, 14). We have also seen that these shared ideas are understood by Rawls as both those of practical reason and the substantive set of liberal principles, because he takes the two to be mutually identifiable (the principles, say of justice as fairness, being the principles and ideas of practical reason passed through the constructivist procedure). In this sense, the principles of justice as fairness reflect ideas of person and society as citizen and fair system of social cooperation respectively. It is these principles that form the content of public reason, the “substantive principles of justice for the basic structure (the political values of justice)” (PL, 253). Alongside these principles are the political virtues that make public reason possible, the ideal that “expresses a willingness to listen to what others have to say and being ready to accept reasonable accommodations or alterations in one’s own view” (PL, 253).

To begin with, the public reason of the political conception allows us to identify what can be removed from the political agenda. Rawls notes that the truth of religious traditions would be removed from the political agenda, because such
truth represents the realm of non-political values around which citizens cannot hope to reach consensus; the appropriate solution instead is to "avoid" issues guaranteed to produce deep disagreement (PL, 151). Rawls suggests that only constitutional essentials and matters of basic justice remain on the political agenda. He also argues that the predominant concern of such core political issues is the acquisition and limitation of political power (cf. PL, 229). Those questions on the agenda then relate to how the principles of justice specify the structures and powers of government, along with a system of "equal basic rights and liberties of citizenship that legislative majorities are to respect" (PL, 228). According to Rawls

[...]his means that political values alone are to settle such fundamental questions as: who has the right to vote, or what religions are to be tolerated, or who is to be assured fair equality of opportunity, or to hold property. (PL, 214)

In this way, according to Rawls, the role of public reason reflects that of the Supreme Court. The Supreme Court serves to protect the "higher law of the people": "[b]y applying public reason the court is to prevent that law from being eroded by the legislation of transient majorities, or more likely, by organised and well-situated narrow interests skilled at getting their way" (PL, 233). Public reason then serves as an interpreter of the principles of justice when applied to core political questions concerning the concretisation of such principles in the governmental structures and systems of rights and liberties of a liberal democratic polity.

Correspondingly, the role of the content of public reason in the search for answers to core political questions is a considerable one. Rawls asserts that "public reason is suitably complete, that is, for at least the great majority of fundamental questions, possibly for all, some combination and balance of political values alone reasonably shows the answer" (PL, 241. Emphasis mine). Public reason in this sense is not understood as offering only a conceptual framework for constraining

58 See also Rawls, "Justice as Fairness: Political not Metaphysical", p. 231.
59 While there are many specifications of government, Rawls notes that the system of basic rights tends to look more or less the same within all constitutional democracies.
the outcomes of public political discourse, rather it appears that Rawls intends the content of public reason *to itself yield solutions* to deep political issues.

To understand why the role and content of public reason are specified in this way, we must remember Rawls's fundamental assertion that justification is addressed to others who disagree with us, and therefore it *must always proceed from some consensus*, that is, from premises that we and others publicly recognize as true; or better, *publicly recognize as acceptable to us* for the purpose of establishing a working agreement on the fundamental questions of political justice. (PL, 229. Emphasis mine)

Agreement must always begin with agreement; without a basis of shared premises there can be no progress on matters of political justice. Hence the *content* of public reason must be a *consensus* on fundamental political principles, and consequently the role of public reason is the deployment of these shared principles for purposes of producing answers on fundamental questions. It is this common understanding that makes further political agreement both possible and likely; pre-established consensus is the prerequisite of all further agreement in a liberal democratic society.

In sum, several points of interest are apparent with regard to Rawlsian public reason. Firstly, public reason again highlights Rawls's interesting interpretation of what constitutes justifiable and legitimate reasons. Again we see that what could be justified toward and accepted by others is equated with that which is likely to be accepted by those others. On this point, Rawls is unequivocal that no "nonpolitical" reasons, i.e. reasons beyond those grounded in the political conception of justice, could possibly, read *would possibly*, be accepted by a plurality holding diverse comprehensive doctrines. Second, the role of the citizen is substantial and vital to the work of public reason; the political virtues require a capacity to *identify, distinguish and translate* between one's private and public values and identity. Third, the content of public reason is tasked with the *production of resolutions* to political disagreements between comprehensive doctrines. Rather than simply offering guidance on how to distinguish via public reason what cannot count as a public reason, it overtakes the discursive political
process by defining its success in relation to its capacity to specify successful outcomes in political disagreement. If we take the role of the citizen (the political virtues) and of the principles of justice (political principles) together, we see that they undertake much if not all of the work of public reason. Very little, it seems, is left for a process of democratic deliberation.60

1.3. Conclusion

Finally, the role of agreement flags the most important concern with regard to this account of public reason. In order for agreement to be possible, public reasoning must begin from shared premises, specifically the political values of the political conception of justice. This premise invites the objection that such shared premises may be, while historically constituted, arbitrary and unjustified, as goes the familiar objection to communitarianism. Rawls attempts to circumvent such an objection by grounding the political values in the shared principles and ideas of practical reason; because the principles of practical reason are those of “common human reason”, they are neither relative nor context-dependent (PL, 55). Also, as the ideas of person and society are complementary ideas of practical reason, these too have some claim to universal validity. The vindication of his starting points therefore rests on the assumption that those to whom justification is owed will not, or cannot reasonably, dispute the basic categories of citizen and well-ordered democratic society. However, it is possible that Rawls’s account of practical reason can be identified as context-dependent; it is a substantive account that does not appear to offer justification for its specific content (i.e. the criterion of reciprocity and the “objective”, “reasonable” citizens’ perspective). As such it may be ill-suited to function as the criterion for justification in a liberal society constituted by an as yet unspecified plurality of ethical perspectives. Rawls loads his account of practical reason with substantive content because he links stability to justification;

---

actual intersubjective endorsement must begin by considering what actual content persons might or would endorse or agree to. Again, I note that this is a separate question to what can or could be endorsed or agreed to.

Furthermore, there are assumptions built into the content of public reason that may raise problems for this account if it turns out that justification is owed to those who fall outside the scope of the overlapping consensus. As noted above, one way this can occur is if persons reject the content of the overlapping consensus as unjustified and context-dependent. A second way this can occur is where persons reject and question their exclusion from the overlapping consensus, i.e. from a particular bounded society. If the ideas of practical reason implicitly assume, and therefore do not justify, certain aspects of liberal democracies, such as the existence of state power or the location of the boundary between those inside and outside a particular liberal polity, then this account of public reason will be unsuited in any context where justification and questions of basic justice cannot be presumed to apply entirely within the borders of a particular liberal democratic polity.

O'Neill has argued that appeals to hypothetical consent suffer from a “deficit of premises” and must therefore introduce unvindicated premises or premises that are only true for some:

To repair the deficit of premises, hypothetical consent theories usually add more specific, empirical premises that refer to features and characteristics that are distinctively true of certain choosers, such as references to their citizenship or membership in certain states or communities. However, in doing this hypothetical consent theories forfeit the possibility of justifying the very features or characteristics they presuppose. Hypothetical consent theories face an uncomfortable choice between idealized (“metaphysical”) justifications and restricted (political) justification. A metaphysical justification fails insofar as it relies on idealized claims that are false of many, even all, choosers; a political justification fails insofar as it relies on conceptions it aims to justify. For example, John Rawls's version of a political justification in Political Liberalism uses conceptions such as those of peoples and citizens in its basic justificatory arguments, so also presupposes boundaries, and thereby also presupposes some exercise of state powers. Appeals to hypothetical
consent that assume and build on political notions such as these cannot justify fundamental political arrangements. (Emphasis mine)

O'Neill suggests that because the particular conceptions of person and society put forth by Rawls implicitly assume the characteristics of citizens and certain features of societies such as the existence of a boundary and the apparatus of state power, neither boundary nor state power receives adequate justification. She also argues that the problem for hypothetical consent, along with actual consent, originates in the attempt to appeal to preference, i.e. to what we might, or actually do, want. Such forms of consent are never sufficient to justify action. O'Neill goes further and argues that it is a misinterpretation to read Kant as endorsing the hypothetical social contract in the constructivist procedure. Instead, she argues that Kant must be read as intending the social contract as a form of possible consent, with an emphasis on the modal verb. Her central claim is that reasoning that justifies action does not need to rely on either hypothetical or actual consent; rather it must focus on what it is possible for agents to accept.

This section has highlighted two important implications for the trajectory of Rawls's reasoning on justice. Firstly, this emphasis on what would be endorsed is not the best (or only) way to discover the appropriate principles of justice because actual acceptance is neither a reliable indicator of justice nor a criterion of justification. Much of the contrivance of Rawls's theory is necessitated by his struggle, on the one hand, to explain how and why actual endorsement is relevant to justification, while, on the other hand, asserting that what is actually endorsed ought not be decisive in the determination of principles of justice.

---

62 “However, the profound limitations of hypothetical consent do not show that theories of justice are impossible. There is no particular reason for thinking that all reasoning that aims to show which types of action (relationship, institution, policy) are acceptable (so setting the context for justifying their particular instances by reference to actual consent) needs to build on a conception of hypothetical consent. The claim that actual consent is a second step, needed to complete a justificatory process (in all but a restricted set of cases, where coercion is shown to be legitimate), rather than a complete justification by itself, could be linked with a variety of conceptions of justification and of justice. However, if the basic moves of a theory of justice do not appeal to consent, we may no longer be considering any recognizable version of the social contract tradition. With these thoughts, back to Kant.” See O'Neill, “Kant and the Social Contract Tradition”, p. 30.
63 O'Neill has written extensively on the dangers of taking actual consent to indicate the justice of institutions or relationships. For some of her most recent thoughts, see O. O'Neill, “Kant and the Social Contract Tradition”, in particular pages 27-30.
Secondly, this struggle leads him to rely on context-dependent criteria, which ultimately compromises both his domestic and international accounts of justice by making his argumentation reliant on *unvindicated and potentially unshareable criteria for reasonableness*. Not only might such criteria rule out the possibility of justifying liberal principles beyond the borders of liberal democracies, but it also might compromise any claim to be viable within a society constituted by a *genuine* plurality of doctrines. Both of these problems can be articulated as a single objection: that his account of justification fails when those to whom justification is owed fall outside the scope of the overlapping consensus.

It was Rawls's contention, however, that the primary context for questions on basic justice and for justification was the "bounded society". "Public" justification, with its criteria of reciprocity and legitimacy, was not owed to those outside the scope of one's particular liberal democratic society. Justice beyond borders was instead defined in terms the state's attitude toward the sovereignty of other nations. These assertions lead him to his particular account of international justice, as set out in *The Law of Peoples*.

Part 2, to which we now turn, addresses the impact of Rawls's early theoretical decisions on his account of international justice. In particular, we will look further at the insufficiency of consent as a fundamental justificatory strategy and the practical implications of this theoretical deficiency in a world of porous and permeable boundaries, a plurality of perspectives on justice and a significant level of underdevelopment and need.
Part 2

*The Law of Peoples*: agreement as the foundation of international law and justice

*The Law of Peoples* was Rawls’s contribution to questions of international justice and stability. In it, he envisioned a Society of Peoples constituted by liberal and “decent” peoples. Decent regimes, while not meeting the conditions of legitimacy specified by a political conception of justice (i.e. the criterion of reciprocity), were nevertheless “acceptable” from the perspective of liberal persons. Though their political institutions denied freedom and equality, their members endorsed a shared conception of nonliberal justice, and on this basis decent regimes were a people entitled to due respect from other societies. It was Rawls’s conjecture that liberal and decent peoples could come to agreement on a set of international principles that would regulate international interaction. Stability and justice would be achieved when all societies were either liberal or decent and agreed to respect the equal standing of others members and abide by the principles of international conduct.

Rawls attempted to merge the practical political reality of international relations with a robust account of global justice that secured the basic rights of the world’s citizens and took seriously the obligations of well-off states to those less well-off. With *The Law of Peoples*, Rawls is not attempting to set out a comprehensive thesis on those individual rights that he believes should be accorded every human being according to their humanity, rather he is attempting to answer what he considered to be a more pressing and practical question; how best can “we”, as Western, liberal nations, engage with other societies and cultures in an effort to forge a
peaceful coexistence, based on mutual respect, between all those nations that treat their citizens with a certain measure of dignity?

While himself affirming a liberal political conception of justice, Rawls argued that stability could not be achieved by attempting to construct an account of international liberal justice. Rather, the ethical pluralism of the domestic society was deepened at the international level, and combined with the rights of “peoples” to determine their own futures, prevented the possibility of extending a full liberal political conception. The fundamental barrier Rawls identified to a just and stable world order was the tension between the demands of liberal justice and the sovereign rights of self-determination of nonliberal peoples. It seemed that to attempt to export liberal political values onto those who rejected them was itself illiberal and, paradoxically, contrary to what he viewed as the liberal values of tolerance and respect.

Rawls believed there could exist societies that rejected liberal political institutions yet sought their own way to honour a domestic conception of justice that concerned itself with the best interests of all members. It was his concern to extend toleration to such societies, which he referred to as “decent”. Such regimes, while not endorsing principles of freedom and equality, did institutionalise consultation procedures that served as a framework within which objection and dissent could be voiced and would be taken seriously. In this way such regimes were fundamentally different from “benevolent dictatorships”, which treated their members well but denied them any form of meaningful participation in the political process, or “rogue states”, which wholly disregarded even “urgent” human rights.

The “acceptability” of decent regimes rests on the consensus or agreement Rawls presumes to exist regarding the conception of nonliberal justice, i.e. that such a conception can be seen as “shared”. By invoking consultation procedures, Rawls is attempting to extend partial legitimation to this agreement. Given that agreements can frequently be spurious, or coerced, or morally questionable, actual agreement requires that some criteria be met in order to grant it legitimacy. In this case “fair”
opportunities for dissent imply that agreement is to some extent free and uncoerced.\footnote{When we address the basis for the shared conception of justice in "decent peoples" we will see how fairness in relation to decent regimes is detached from the criterion of reciprocity and linked to opportunities for dissent.} On this basis decent societies can be seen as peoples who share a conception of justice, as opposed to states that impose a political order on their members.

While much of what Rawls aims for is admirable and well-intentioned, I argue that the tension between respect and justice that he identifies as a barrier to a just international order is in fact the result of his inability to justify his own conception of justice beyond the boundaries of a particular liberal society. The direction of my critique is therefore to analyse his fundamental justificatory strategies before critiquing his conceptions of international justice and “decency” in light of the deficiencies I identify. I will examine how the origins of the problem lie in Rawls’s early theoretical decisions, as was discussed in Part 1. In particular, the distinction between probable over possible consent will be discussed in relation to O’Neill’s arguments for the conditions for the possibility of consent or agreement.

The critique begins by considering the grounds for claims of impartiality or objectivity, or, as discussed by Forst, the possibility of claiming a “universal-moral” basis of justification in Rawls’s account (2.1.). I argue that, by making the justification of his fundamental conceptions of practical reason reliant on what the audience of justification may affirm, Rawls forsakes any claim to the impartial or objective validity of his account of justice. In particular, this has implications for the capacity of this account to distinguish between reasonable and unreasonable perspectives. This critique argues that such a distinction could only be made by an account that was independent of the perspective of those who already agree on the content of what is reasonable. What will also be addressed is Rawls’s understanding of legitimacy, as this understanding plays a central role in explaining why Rawls takes the (hypothesised) actual acceptance of decent regimes to ground an assumption of at least partial legitimacy. Legitimacy cannot be separated from moral justifiability, and Rawls’s attempt to do so belies the problematic role of consent in his justificatory strategies.
Having critiqued what I describe as the insufficiency of consent as justification for its own conditions, I turn to *The Law of Peoples* to explore the implications Rawls's deficient justification has on his understanding of international justice (2.2.).

The focus of the critique will be on the account of decent regimes and the agreements reached in the international original positions. Of decent regimes, the primary objection relates to the inconsistencies of this account. In particular I highlight the inconsistency of describing decent regimes as a “people” when other features of such regimes rule out the possibility of assuming the necessary level of homogeneity with regard to the conception of justice in the public sphere of such regimes. Rather, a level of plurality exists that undermines the apparently “shared” nature of the conception of justice. I will argue, based on these inconsistencies, that the *affirmation* of a nonliberal conception of justice lacks a defensible form of legitimacy. Furthermore, *without* such legitimacy, the grounding of a concept of a “people” is in jeopardy. This deficit has serious implications for Rawls's account of international justice.

Rawls's deficient justificatory strategies also have implications for the role of boundaries in his account of international justice (2.3.). Consensus alone cannot justify, since a consensus on morally questionable goals is thinkable. Boundaries, which in Rawls's account of domestic justice coincide with the limits of the overlapping consensus, i.e. with the particular political conception of justice, cannot be taken as assumptions prior to justification. The justification for boundaries must extend *and be extendable* to those outside the consensus, whether excluded as foreign or as unreasonable. I argue, based on the cosmopolitanist critiques of Charles R. Beitz and Thomas Pogge, that the bounded state cannot serve as the horizon of considerations of justice. For them, those outside the consensus are also entitled to justification, since they are relevant others entitled (also) to justice. In order to examine the grounds for delineating the scope of moral concern, I turn to O'Neill's defence of a cosmopolitan perspective on justice (2.4). I will conclude that the duty of assistance is a wholly inadequate conceptual device for understanding the nature of transnational
justice, as the concept of assistance turns out to deny that we have real obligations of justice toward distant others. Ultimately, Rawls's account of international justice relies too heavily on the perspective of the citizens of liberal states and fails to produce an account that can be justified as to all persons to whom it is relevant.

2.1. The conception of reasonableness as an “idea of practical reason”: deficiencies in Rawls's fundamental justificatory strategy

As was noted in the introduction, Rawls treats the apparent affirmation of a common good idea of justice in decent regimes as an indicator of (at least partial) legitimacy. I argue that his implicit assumption of the legitimacy of an actual agreement in decent regimes (as opposed to the conditions that make agreement legitimate) is the result of his fundamental justificatory strategy. In order to do so I begin by analysing the fundamental justification Rawls offers for the conceptions of practical reason, in particular of reasonableness, that ground the apparent "objectivity" of the citizens' perspective in Political Liberalism, and question whether his strategy can be defended as non-biased and impartial (2.1.1).

In Part 1 we noted that Rawls conflated possible and probable consent or agreement, with a psychologising emphasis on what would be agreed to as opposed to what could (in a morally justified sense) be agreed to. Here it is emphasised that the salient distinction between possible and probable consent is that there is no guarantee that probable consent meets the conditions for the moral justifiability or legitimacy of consent.

I revisit Forst's comments on the possibility of a universal-moral level of justification in Rawls's account of justice, specifically the claims that the principles of political liberalism were those principles that were necessary in democratic states, and that Rawls's account of practical reason was an objective account that "could not be reasonably rejected" (CJ, 174). In Part 1 I concluded that Rawls's account of practical reason stood in need of further justification, and cannot be accepted as objective without an adequate account of why it could not be
reasonably rejected by “reasonable” persons, i.e. what gave it objective validity. It is now apparent, from Rawls’s comments in *The Law of Peoples*, that his account of practical reason is ultimately justified by reference to what “reasonable” persons will endorse or affirm on due reflection.

Rawls’s account of practical reason is justified in the end via a consensus or agreement on its content. For this agreement to ground objectivity, it would have to be demonstrated how and why it is a consensus of the “reasonable”. I argue that this objectivity is unavailable, because the concept and content of reasonableness has its origin in a specific understanding of practical reason in which it is justified by consensus, by *actual* agreement.

With these considerations in mind, I turn to look at Rawls’s understanding of legitimacy (2.1.2.). I review O’Neill’s critique of consent as justification, in particular her conclusion that neither actual nor hypothetical consent can account sufficiently for the legitimacy of consent. She argues that consent is a form of agency, and as such cannot be conceived as legitimate where the underlying capacities for agency are lacking. I also note that neither actual nor hypothetical forms of consent can justify the conditions for possible consent. I conclude that, in opposition to Rawls’s understanding, legitimacy cannot be separated from the conditions for the possibility of consent or agreement.

While actual agreement is central to legitimacy on Rawls’s understanding, as we have noted it is also central to his understanding of moral justification, as the conception of reasonableness that underpins the justifiability of agreement in political liberalism is itself reliant on affirmation.

I conclude by reiterating that it is not sufficient to designate those outside such a consensus as unreasonable. Such a claim could only be valid if it was grounded in an argued and nonpartisan perspective *independent* of the consensus of those who designate themselves as reasonable. I note that a properly impartial perspective, such as that of O’Neill, would be capable of demonstrating why certain conditions
are necessary, as opposed to only preferential for certain persons, and in this way would be capable of justifying the moral-universal account of practical reason rather than presupposing a restricted political account.

2.1.1. A consensus of the “reasonable”: the justification of Rawls’s conception of practical reason

In *Political Liberalism*, Rawls considered what conception of justice would be agreed to by a plurality of ethical doctrines. He conjectures that only principles meeting the criterion of reciprocity would be acceptable to such a plurality. In this way he links probable agreement or consent to (a version of) possible agreement or consent; morally acceptable principles such as those of justice as fairness are also those that are likely to be accepted by a reasonable plurality of ethical doctrines. Rather than taking agreement or acceptance of any principle as legitimate, Rawls insisted that only a reasonable agreement, one that endorsed the principle of reciprocity, could be taken as defensible. The agreement of citizens, hence the overlapping consensus and public reason, could be defended so long as the principle of reciprocity grounded the agreement. In decent regimes, however, actual agreement has been decoupled from moral principles such as reciprocity, and yet is still taken to count toward legitimacy. This is, I suggest, a result of the difficulties surrounding Rawls’s account of practical reason (and justification).

In Part 1, we examined Rainer Forst’s critique of both Rawls and some of his critics. Forst analysed the suggestion that the use of conceptions of citizen and liberal democratic society in Rawls’s work implied that his account was contextual. Forst argued that these conceptions were articulations of conceptions of person and society intended as complementary ideas to Rawls’s account of practical reason, and that this account was intended by Rawls to be non-contextual and hence could ground a claim to universal-moral justification. The principles of practical reason defined the condition of reasonableness that grounded the

---

65 “Necessity” refers to practical considerations regarding the conditions for the possibility of action in relation to finite and vulnerable agents, i.e. principles that support agency and reject its violations.
citizens' perspective and were not reasonably rejectable in a moral sense. Rather than merely articulate a particular set of values for a particular society, the principles set out what was necessary for a democratic society to be grounded in shareable principles. Given that these principles were available to all moral persons (or "latent in common sense"), though not necessarily endorsed by all persons, the objective perspective was also therefore not restricted to the point of view of historically situated citizens of liberal societies. Hence, public justification through the overlapping consensus in Political Liberalism could not be considered contextual because the citizens' perspective that grounded the justification was itself grounded in an account of practical reason that had some claim to universal validity i.e. in a more foundational and "freestanding" level of justification.

I suggested, however, based on O'Neill's critique, that Rawls's understanding of practical reason was itself problematic. It was a substantive and content-based account that defined itself by a set of principles that it did not appear to offer justification for.66 These unvindicated principles were inextricably linked to conceptions of citizen and liberal society and, as a result, I questioned whether there was in fact an objective basis for the account. Without a satisfactory vindication for the derivation of principles of practical reason, I concluded that Rawls could not escape the charge of contextualism; his account of practical reason could not function as a basis of justification as it stood itself in need of justification.

In The Law of Peoples Rawls makes explicit the derivation of the principles of practical reason:

I should say the following: at no point are we deducing the principles of right and justice, or decency, or the principles of rationality, from a conception of practical reason in the background. Rather, we are giving content to an idea of practical reason and three of its component parts, the ideas of reasonableness, decency, and rationality. The criteria for these normative ideas are not

---

66 The focus of this critique is the problematic justification offered by Rawls for his particular restricted and political account of practical reason, as put forward in Political Liberalism and The Law of Peoples. I do not here address the adequacy of such a conception of practical reason. In Part 3, however, we will examine how O'Neill offers an account of Kantian practical reason that is not reduced or limited to the political domain of liberal democracies, but can function as a fundamental and justified form of moral reasoning accessible to all persons due to their inherent moral capacity.
deduced, but enumerated and characterized in each case. Practical reasoning as such is simply reasoning about what to do, or reasoning about what institutions and policies are reasonable, decent, or rational, and why. There is no list of necessary and sufficient conditions for each of these three ideas, and differences of opinion are to be expected. We do conjecture, however, that, if the content of reasonableness, decency, and rationality is laid out properly, the resulting principles and standards of right and justice will hang together and will be affirmed by us on due reflection. Yet there can be no guarantee. (LP, 86-7. Emphases mine)

The derivation of the account of practical reason can then be explained as follows. The content of the relevant principles is “properly” set out, in the sense that their criteria are “enumerated and characterised”. Their justification is then held back until two conditions are met. Firstly, the principles of justice that emerge must fit together coherently, and, secondly, they must then be “affirmed by us on due reflection”. Two points are clear: first, the content of principles of practical reason, e.g. the conception of reasonableness, are enumerated rather than derived from a more fundamental account of reasoning, and, second, the vindication for this set of criteria is that they are endorsed or affirmed “by us”, the audience of justification. Given that The Law of Peoples is directed only at those who begin from the perspective of the political liberal conception of justice, we may assume that the audience of justification is restricted to citizens of liberal democratic societies. Endorsement or affirmation, or what I would also refer to as acceptance, must therefore be understood to function at the most foundational theory level; not only does it justify the principles of justice on due reflection, it also appears to justify the account of practical reason itself.

67 While the perspective of decent peoples is considered, it is done so as a form of hypothetical speculation on the part of liberal persons, who remain the audience of justification: “Finally, it is important to see that the Law of Peoples is developed within political liberalism and is an extension of a liberal conception of justice for a domestic regime to a Society of Peoples. I emphasize that, in developing The Law of Peoples within a liberal conception of justice, we work out the ideals and principles of the foreign policy of a reasonably just liberal people... The reason we go on to consider the point of view of decent peoples is not to prescribe principles of justice for them, but to assure ourselves that the ideals and principles of the foreign policy of a liberal people are also reasonable from a decent nonliberal point of view. The need for such assurance is a feature inherent in the liberal conception. The Law of Peoples holds that decent nonliberal points of view exist, and that the question of how far nonliberal peoples are to be tolerated is an essential question of liberal foreign policy” (LP, 10. Emphases mine). What is notable from the above passage is that “we” are clearly those who share in the ideals of liberal political justice, while the members of nonliberal societies are referred to as “them”, i.e. the other, and not the audience of justification.
In Political Liberalism, full justification, where citizens endorse the principles of justice by embedding the political conception in their comprehensive doctrines, and public justification, where they collectively endorse the principles and hence form an overlapping consensus that defines the account of public reason, were separate and subsequent to pro tanto justification, which offered an internal justification for the political conception of justice. If one interprets pro tanto justification as attempting to offer a self-contained “moral” justification for political liberalism, as Forst suggests in Contexts of Justice, the reflective equilibrium at the full and public levels of justification could be understood as grounded in an understanding of practical reason that is prior to it.

This is in line with Forst’s claim that Rawls’s account did not attempt to merely articulate what was present in a particular culture, but attempted to give an account of what must necessarily be present in order for any society to be based on shareable principles. However, Forst critiques Rawls on the grounds that he is unable to decide finally whether the more fundamental justification is that of “freestanding”, “moral” justification, or whether it is “ethical-comprehensive”:

This account raises the central question of how the moral force that the freestanding pro tanto justification of the conception of justice initially confers entirely independent of ethical beliefs can, on the one hand, be absorbed wholly by the ethical “truth” of comprehensive doctrines, while, on the other hand, it prevails justice in the political-public use of reason in its restriction to moral-political values of justice. Rawls is unable clearly to explain the moral justification of the political conception: he fluctuates between a form of justification based on an ethical-comprehensive doctrine and a freestanding moral justification. But ultimately he must opt for the latter, since otherwise the first type of justification would fail... The level of justification that is reached at the first step in a public, reciprocal and general justification must govern the other steps, for otherwise there could be no insight at all into the priority of justice over “non-political” values.60

Rawls does not make clear whether the ultimate justification is universal-moral or ethical-comprehensive, and Forst argues that this deficiency jeopardises both forms of justification by undermining the priority of justice, and that only a “public, reciprocal and general justification” could serve as adequate justification.

---

would suggest that Forst is correct to question the universal validity of the “freestanding” form of justification, and I argue that Rawls’s comments on the derivation of practical reason indicate his account is not fundamentally reliant on universal-moral arguments. However one interprets the relationship between pro tanto, full and public justification, what remains apparent is that the procedure of endorsement of principles through the overlapping consensus is reliant on Rawls’s account of practical reason, in particular a conception of reasonableness, to ground its deployment as a form of justification, or as an “objective perspective”. This may initially appear to support the claim that the endorsement levels of justification are grounded in a prior level, which receives independent moral justification. However, Rawls’s above statement from The Law of Peoples regarding the justification and derivation of practical reason returns us to our earlier consideration that this account of practical reason, in particular the conception of the “reasonable”, lacks adequate justification. What is now made explicit is that his understanding of practical reason is justified via the endorsement of the audience of justification. It is now apparent that endorsement, or acceptance, or a consensus on the content of a particular conception of reasonableness, is fundamental not only to the justification of principles of justice, but also to the account of practical reason itself.

Yet, as we have discussed, it must be shown why an agreement around certain standards or criteria is not morally arbitrary, i.e. why the affirmation of a particular conception, such as reasonableness, can function as justification. Agreement on its own is no indicator of moral significance. It is at this juncture that the limitations or deficiencies of Rawls’s justificatory strategy become apparent. In Political Liberalism, objectivity is grounded in consensus; the objective perspective is given by citizens’ cumulative endorsement of the political conception of justice. The overlapping consensus is therefore the public basis of justification. This consensus, this endorsement of a certain conception of justice, could function as justification (as the objective perspective) because the perspective of those who formed the consensus that defined the objective perspective was the perspective of persons who were understood to be “reasonable”. Unreasonable persons were
excluded from the consensus and their agreement on principles could not be taken to reflect any form of justification or legitimacy.

Reasonableness, therefore, is a necessary condition that must be met in order for a particular perspective to qualify as objective. However, as we have just seen, the content of reasonableness on Rawls’s account, as an “idea of practical reason”, is justified by consensus or agreement on its content. Its only apparent vindication is that, if it is set out “properly”, the resulting principles are “affirmed by us on due reflection”. According to Rawls, endorsement or consensus is objective when it is that of “reasonable” persons, however what constitutes reasonableness appears also to be grounded in endorsement; reasonableness and consensus appear to be mutually reliant upon each other for their objectivity.

Rawls claims of course that the perspective of liberal citizens is the only available starting point, that reason must be that of partially contextualised agents, because reason must always be “from somewhere” (PL, 116). However, it is clear that this perspective relies on an understanding of reasonableness that, it now appears, cannot be justified as a morally necessary conception. There is a circularity to his strategy of justification; what he claims as the ultimate ground of justification, the perspective of moral (reasonable) persons, itself takes for granted the premises and standards that it seeks to justify, e.g. reasonableness. In other words, Rawls relies on actual endorsement to ground the objective criteria that apparently make the consensus legitimate.

The problem is that Rawls gives no account of why a consensus around this account of practical reason is objectively morally superior or more legitimate than a consensus around a different set of principles. Rather, the circularity of the justification implies that the conception of the reasonable that is intended to ground the citizens’ perspective as “objective” is itself dependent on that perspective for its own justification.

In this way we see why Forst’s comments regarding Rawls’s understanding of practical reason do not quash the objection that his thesis is grounded in a
context-dependent account of practical reason. Forst suggested that Rawls’s justification strategy could be grounded in an account of practical reason that could not be reasonably rejected in a moral sense; it described the necessary conditions for shareable principles. What is now clear is that, in contrast to Forst, Rawls was not considering the necessary conditions for the possibility of sharing principles in an abstract, universal-moral, sense, but rather the conditions for the probability of sharing principles in a particular society. Rather than considering what could be universally shared, Rawls focused his justification on whether certain conceptions of practical reason were in fact shared. The vital distinction is that whether or not an actual agreement will be formed around a particular set of principles is not an indication of whether those principles are morally justified or fundamentally shareable. Rawls does not offer an account of what makes his understanding of practical reason morally necessary, as opposed to only preferential from a particular perspective, despite the fact that it is the apparent justified objectivity of this account that grounds the agreement or consensus that is so fundamental to his account of justice.

2.1.2. Clarifications made in debate with other conceptions: Rawls’s discussion of legitimacy and of actual and hypothetical agreement as forms of fundamental justification

The “acceptability” of decent regimes rests on the consensus or agreement Rawls presumes to exist in these societies regarding the conception of nonliberal justice. Specifically it appears, as we will see when we examine the role of loyalty and affirmation in decent regimes, that Rawls takes the actual affirmation of the conception of nonliberal justice to ground a claim to the legitimacy of the political institutions of decent regimes. In what follows I will review his understanding of legitimacy and argue that it is not possible to separate the validity of agreement from the consideration of its moral justifiability, based on O’Neill’s account of consent as a form of agency that therefore requires capacities for agency.
In his "Response to Habermas", Rawls sets out his understanding of legitimacy, in distinction from Habermas, who he sees as mistakenly equating "legitimate" with "just":

To focus on legitimacy rather than justice may seem like a minor point, as we may think 'legitimate' and 'just' the same. A little reflection shows they are not. A legitimate king or queen may rule by just and effective government, but then they may not; and certainly not necessarily justly even though legitimately. Their being legitimate says something about their pedigree: how they came to their office. It refers to whether they were the legitimate heir to the throne in accordance with the established rules and traditions of, for example, the English or the French crown.69

Legitimation is at first tied to sets of "rules" and "traditions", and whether a particular ruler or government came to office "in accordance" with these rules or traditions. For democratic legitimacy, Rawls makes a similar claim by tying the legitimacy of particular legal acts or statutes to their being in line with the framework specified by the formation of the constitution:

The same holds under a democratic regime. It may be legitimate and in line with long tradition originating when its constitution was first endorsed by the electorate (the people) in a special ratifying convention. Yet it may not be very just, or hardly so; and similarly for its laws and policies. Laws passed by solid majorities are counted legitimate, even though many protest and correctly judge them unjust or otherwise wrong.70

69 John Rawls, "Political Liberalism: Reply to Habermas", The Journal of Philosophy, Vol. 92, No. 3 (March, 1995), pp. 132-180, p. 175. Habermas, on the other hand, understands the concept of legitimacy in Kantian terms as contrasting with the concept of mere "legality", and suggests that legitimacy cannot be detached from conditions of equal liberty, on the basis that citizens must also be capable of obeying the system of law to which they are subject, and this presupposes subjective liberty rights: "According to Kant's conception of legality, coercive law extends only to the external relations among persons and addresses the freedom of choice of subjects who are allowed to follow their own conception of the good. Hence modern law, on the one hand, constitutes the status of legal subjects in terms of actionable subjective liberties that may be exercised by each according to her own preferences. Since it must also be possible to obey a legal order for moral reasons, the status of private legal subjects is legitimately determined by the right to equal subjective liberties. As positive or codified law, on the other hand, this medium calls for a political legislator, where the legitimacy of legislation is accounted for by a democratic procedure that secures the autonomy of the citizens. Citizens are politically autonomous only if they can view themselves jointly as authors of the laws to which they are subject as individual addressees." Habermas, "Reconciliation Through the Public Use of Reason", p. 130. Such a position could be compared to O'Neill's position that forms the basis for our critique, which argues, for example as we have seen in "Kant and the Social Contract Position", for the necessity of constitutional principles of freedom and equality in order to ground the possibility of universal consent.

70 Rawls, "Political Liberalism: Reply to Habermas", p. 175.
What is decided by members of a democratic society in accordance with an “endorsed” constitution is legitimate, though not necessarily just, because there can be no guarantee that what is agreed to by a majority is always just. However, it is prevented from being deeply unjust by being in accordance with a constitution that was at some point endorsed:

Democratic decisions and laws are legitimate, not because they are just but because they are legitimately enacted in accordance with an accepted democratic procedure. It is of great importance that the constitution specifying the procedure be sufficiently just, even though not perfectly just, as no human institution can be that. But it may not be just and still be legitimate, provided it is just enough in view of the circumstances and social conditions. A legitimate procedure gives rise to legitimate laws and policies made in accordance with it; and legitimate procedures may be customary, long established, and accepted as such.

Legitimacy attaches to what is actually agreed within processes whose structure and framework have also been agreed. Justice and moral justifiability, on the other hand, relate to whether conditions of freedom and equality have been established as constraints on what can be agreed. This distinction allows Rawls to claim that decent regimes can be at least partially legitimate while not grounding their political procedures in principles of freedom and equality. Actual agreement, and not morally justifiable agreement, is the object of legitimacy.

While it is not possible to attempt a full examination of why an agreement that may not meet standards of justice can or should be perceived as legitimate or valid, it is important to note that Rawls correctly understands the need to account for the status of actual agreements where full conditions of justice are missing. It may not be acceptable to reject all such agreements in such circumstances, rather the decisions made by persons, understood as agents, must be taken seriously. However, while it may not be acceptable to override actual agreements where conditions of justice are lacking, I argue, based on O’Neill’s Kantian account of the conditions for possible consent or agreement understood as a form of agency, that

71 Rawls, “Political Liberalism: Reply to Habermas”, p. 175.
it is also not possible to separate the legitimacy of agreements from their moral justifiability.\footnote{There remains a question on the status of actual agreement. While I cannot attempt a full analysis of this question, one important consideration may be that we must distinguish the illegitimacy of an agreement (from the perspective of moral justifiability) from the right to override such an agreement. Actual agreements in conditions that lack freedom and equality may not be legitimate, but this does not imply that they may also be disregarded or invalidated by other agents, for example, in the case of decent regimes by external actors. I would suggest in the place of Rawls’s position of non-intervention that supports the status quo, based on O'Neill’s account of justice, that the status of agreements as illegitimate instead points to the need to institutionalize support for the capacities for agency that ground the legitimacy of agreement. Overriding such agreements may only serve to further undermine capacities for agency and would therefore be incompatible with the idea of moral justifiability that underpins legitimacy.}

To understand the problematic implications of Rawls’s understanding of legitimacy I will return to O’Neill’s critique of social contract theories in her article “Kant and the Social Contract Tradition”. O’Neill argues that while the social contract tradition has held to the view that some kind of consent, either hypothetical or actual, can justify political arrangements, the problems of both accounts are sufficient to undermine the claim that consent itself is fundamental to justification or legitimacy. With regard to actual consent, she argues that “[t]here are some cases where actual consent is not sufficient to justify and others where it is not necessary”. With regard to actual consent being insufficient to justify:

Suppose that A agrees that B may practice surgery on him... or to let B bully him, or to be B’s slave. In these... cases we are likely to think that, all legalities aside, consent does not provide sufficient justification. We may be tempted to “save” the claim that actual consent is necessary for justification by arguing that in such cases there was something defective about the consent. We may argue, for example, that the danger or plain stupidity of consenting to dangerous treatment or the profound violation of one’s rights is evidence that such action cannot have been genuinely consented to by competent adults, that here consent is defective, just as in other cases it is defective because the consenting parties were not adequately rational, informed, free from duress, and the like. However, this is a desperate line of argument: there is all too much evidence that people sometimes genuinely consent to action that seems deeply unacceptable, even to action that profoundly injures, oppresses or degrades them. Across the board insistence that any consent to such action must be flawed merely suggests an underlying refusal to consider the possibility that justification requires more than actual consent.\footnote{O’Neill, “Kant and the Social Contract Tradition”, p. 30. Further references in text (KSCT)} (Emphasis mine)
Actual consent in many cases seems insufficient to legitimise certain types of activity, in particular those that oppress or degrade others. O’Neill also cites examples where consent is not necessary to justify action, such as when “laws are enforced” or “when public authorities take emergency action” and concludes that that “actual consent is not always needed: it is no more necessary than it is sufficient to justify action” (KSCT, 28). Hence, O’Neill suggests that justification may in fact have “two stages”:

First a certain type of action (relationship, institution, policy) must be shown acceptable; second, the parties involved in or affected by a particular instance of such an action (relationship, institution, policy) must (generally) consent to it... Daily life is full of examples of two-stage justifications. The myriad activities and transactions of domestic, commercial, and professional life can be justified by showing that they are consensual, provided that such activities and transactions are of acceptable types, but not otherwise. Political life incorporates processes for allowing for actual consent and dissent... provided that they are of acceptable types. (KSCT, 28)

Finally, she notes that actual consent is a propositional attitude and therefore also referentially opaque, hence “actual consent will not automatically transfer from its initial object to the logical or causal implications of that initial object” (KSCT, 29). There are many important aspects of proposed acts or courses of action which cannot be automatically presumed to be consented to by the agent, because these may be obscured for the agent.

Fundamental to this argumentation is that O’Neill emphasises the distinction between possible and probable consent. This can now be reformulated as the claim that there are considerations prior to the stage of consent (whether actual or hypothetical) that are decisive in terms of the justifications of actions, policies, political institutions etc. O’Neill notes that Kant placed emphasis on the modal verb in his “most explicit passage on the social contract”. This, she claims, indicates that for Kant only “the criterion of possible consent can vindicate a republican constitution that guarantees freedoms within the law: constitutions that lack this structure cannot, he claims, be universally consented to” (KSCT, 33).
While a fuller examination of the foundations of O’Neill’s Kantian position will have to wait until Part 3, for now I will briefly review her account of his justification of the republican constitution, as it demonstrates the crucial distinction between the justificatory strategies of possible and probable consent, and will allow us to return to Rawls’s steps of argumentation. O’Neill’s claim is that Kant justified the “basic components of a legitimate republican constitution” by arguing that they were necessary in order to secure the possibility of universal consent:

If the idea of a social contract is that of a constitution that could secure universal consent, then any constitution that exemplifies it must require the freedom of individuals, without which the possibility of genuine consent or dissent is undermined, at least for some, and universal consent becomes impossible. Second, it requires their common dependence on or subordination to law: if anyone was above or outside the law, freedom could be systematically or gratuitously undercut, and once again the possibility of genuine consent or dissent is undermined, at least for some, and universal consent becomes impossible. Third, it must endorse the legal equality of citizens, since the subordination of some individuals to others (rather than to the law) would once again undercut freedom. And with it the possibility of genuine consent or dissent, at least for some, and universal consent becomes impossible. In saying that these principles must be exemplified in any constitution that can be derived from the idea of the original contract, Kant does not insist on constitutional uniformity, but he does claim that all constitutions must meet these three quite demanding “republican” conditions. (KSCT, 33)

Here we see that the problem with claiming legitimacy for institutions that deny freedom and equality (such as those of decent regimes) is that certain persons are denied the requisite equal freedom required for their consent to be considered possible at all. Where freedom and equality are not guaranteed, the possibility of free and uncoerced consent is undermined and any form of apparent agreement, consensus (e.g. in the form of loyalty) cannot be presumed to be (even partially) legitimate. O’Neill’s crucial observation is that there are certain conditions that must be met for consent to be considered legitimate, and an obvious corollary of such a consideration is that consent cannot be deployed to legitimate its own conditions or criteria.
Rawls does not think actual consent alone an appropriate instrument for identifying or justifying political institutions or universal moral duties such as justice. This is emphasised by his assertion that obligations (which result from voluntary acts) cannot arise in relation to unjust institutions because the "necessary background does not exist for obligations to arise from consensual or other acts, however expressed" (TJ, 112), rather the first part of the principle of fairness "formulates the conditions necessary if these voluntary acts are to give rise to obligations" (TJ, 112). Furthermore, he comments that it is a mistake to argue against justice as fairness and contract theories generally that they have the consequence that citizens are under obligation to unjust regimes which coerce their consent or win their tacit acquiescence in more refined ways. Locke especially has been the object of this mistaken criticism which overlooks the necessity for certain background conditions. (TJ, 112. Emphasis mine)

Thus, in his view, actual consent is not guaranteed to incorporate the relevant just background considerations that lend legitimacy to consent and is therefore not an appropriate form of consent for justifying arrangements.

Rawls’s response to these considerations is not, however, to question the status of consent as fundamental to justification. Instead he turns to hypothetical consent, which, he believes, in contrast to actual consent, is perfectly capable of modelling the appropriate considerations, for example in the original position, and can therefore resolve the shortcomings of actual consent and demonstrate the moral justifiability of the agreements conceived in the hypothetical scenario. In the case of justice as fairness, hypothetical consent is able to argue that both unconditional natural duties and a principle of fairness would be accepted under the right conditions.74

74 With regard to natural duties, Rawls argues that without them “the public conviction that all are tied to just arrangements would be less firm, and a greater reliance on the coercive powers of the sovereign might be necessary to achieve stability. But there is no reason to run these risks. Therefore the parties in the original position do best when they acknowledge the natural duty of justice. Given the value of a public and effective sense of justice, it is important that the principle defining the duties of individuals be simple and clear, and that it ensure the stability of just arrangements. I assume, then, that the natural duty of justice would be agreed to rather than a principle of utility, and that from the standpoint of a theory of justice, it is the fundamental requirement for all individuals” (TJ, 337). And with regard to the principle of fairness/obligation he explains why “having trust and confidence in one another, men can use their public acceptance of
In this way, hypothetical consent shares a good deal in common with considerations of possible consent. Both recognise that actual consent cannot on its own function as sufficient or necessary justification, and that certain criteria must be met before actual consent and consent transactions become relevant to justification, as in the two-stage process suggested by O'Neill. By turning to hypothetical consent Rawls recognises that there is a prior stage to actual consent that addresses the conditions that frame consent transactions. However, while both recognise this necessary preceding stage, an examination of their strategies of justification belies the significant disadvantages of the hypothetical consent procedure, and the corresponding advantages of turning to consider possible consent, i.e. of recognising that consent, whether actual or hypothetical, may not be fundamental to justification.

In order to circumvent the difficulties of actual consent, the original position hypothesises a scenario whereby the normal conditions that undermine the legitimacy of consent (by encouraging agreements around unfair principles) are bracketed. While this is commonly thought of as a process of abstraction (notably by Rawls) O'Neill has argued in several of her works that it is in fact a form of idealisation. When discussing the justification of *A Theory of Justice*, she comments that

[abstraction, taken strictly, is simply a matter of leaving open whether or not certain predicates are true of the matter at hand. In abstracting from available starting points we may omit assumptions from which we began, but we will not introduce assumptions with which we did not

these principles enormously to extend the scope and value of mutually advantageous schemes of cooperation. From the standpoint of the original position, then, it is clearly rational for the parties to agree to principles of fairness... At the same time, given the principle of fairness, we see why there should exist the practice of promising as a way of freely establishing an obligation when this is to the mutual advantage of both parties. Such an arrangement is obviously in the common interest. I shall suppose that these considerations are sufficient to argue for the principle of fairness" ([TJ], 348). While it is not possible to fully review these arguments here, it is important to note that grounding natural duties in conditions of reciprocity seems incompatible with the notion of "unconditionality", which suggests an asymmetrical agent-centred perspective on duty such as that of O'Neill's Kantian account. The distinction between duty and obligation is also not employed by O'Neill, rather her analysis of principles of justice treats obligations as equivalent to duties in that their derivation is not attached to voluntary acts but to practical considerations about the conditions for just interaction.
begin. Abstraction in this strict sense is a (relatively) safe method of argument because it does not introduce additional, possibly questionable or false, materials for reaching conclusions.^

The crucial point about abstraction is that it does not, and must not, introduce "additional, possibly questionable or false" premises that are to apply to a given situation. In this way abstraction rests on fairly uncontroversial grounds. This is not however the strategy deployed by Rawls in constructing the original position:

Rawls's notion of the 'original position' is evidently no mere abstraction, although it is abstract in many respects. The defining ignorance of those in the original position is not the abstract claim that they may or may not have information of various sorts, but the idealized claim that they are without certain sorts of information... Within the conception of the original position, in which rational self-interested persons choose behind a veil of ignorance, there is much that is no mere abstraction.^

The introduction through the veil of ignorance of conditions that bracket natural and social contingency, i.e. that bracket those conditions deemed unfair or arbitrary from the perspective of justice, is an idealisation of actual situations. This, she concludes, undermines the foundational assumptions of the theory and calls its justification into question:

Once we begin to consider the numerous ways in which Rawls uses not only the uncontroversial strategy of abstraction, but the controversial strategy of idealization, the status of many of the assumptions built into the original position, and more generally into his method of justifying principles of justice in *A Theory of Justice*, may seemingly be questioned.... Idealization is always a questionable method because it may introduce false premises and consequently may lead to false conclusions. *It is acceptable only where each idealization is vindicated.* Yet many of the idealizations Rawls incorporates into his original position are not vindicated. At the very least a further justificatory strategy is needed.^

The original position hypothesises an idealised scenario that deploys idealised assumptions about the knowledge and motivations of human agents in order to argue for principles of justice that are to apply in the nonideal world. As O'Neill

---

points out, "[w]hy should we accept principles as fundamental to justice on the
basis of the claim that they can be generated by curious hypothetical
procedures?" 78 Whereas abstract premises deploy only fairly uncontroversial
assumptions 79 and can therefore justify themselves as relevant to actual persons,
in order for hypothetical consent methods to demonstrate such relevance they
must show why their particular idealisations should or ought to be applicable to
actual persons.

As we noted earlier, hypothetical consent seems to face a choice between two
unsatisfactory options; either deploying idealised conceptions that are irrelevant
to actual persons, or else attempting to "repair the deficit of premises" by adding
premises that are true of only certain persons, thereby forsaking non-
contextualised relevance (KSCT, 29-30). O'Neill suggests as an example that in
presupposing peoples and boundaries, Rawls fails to justify these concepts. She
concludes again that consent, in this case hypothetical consent, is not sufficient to
demonstrate the permissibility of certain types of action (cf. KSCT, 30).

As we have seen, actual consent is neither necessary nor sufficient to justify
arrangements, principles or institutions. Hypothetical consent also cannot by itself
explain why the criteria of its construction are necessary or relevant to justice, and
struggles also to vindicate its premises. While it can attempt to model conditions
for actual consent it must introduce questionable assumptions that undermine the
agreement that emerges.

Neither the hypothetical nor the actual form of consent function as the basic level
of justification, the fundamental deficiency remains; consent, of either kind, cannot
function as the basis of justification. As O'Neill has argued, consent is the second
consideration in a two-stage process. The first stage, which we noted that Rawls
himself acknowledges, involves meeting certain criteria, in other words it involves

79 Such fairly uncontroversial assumptions include, for example, limited and varying capacities for
rationality and agency in humans, along with limited and varying measures of dependence,
independence and interdependence. It is these minimal assumptions that underpin O'Neill's
alternative Kantian interpretation, which will be discussed in Part 3.
vindicating the conditions that legitimise consent. Despite Rawls's recognition of the importance of fair background conditions for consent, he fails to interpret this as the insight that consent itself is not fundamental to justification.

Therefore, despite his misgivings, actual agreement is the indicator of legitimacy in Rawls's account. Furthermore, with regard to justified agreements, the conception of reasonableness that is needed to ground the objectivity of the perspective of those who agree is, as we have seen, too reliant on that perspective to ground the apparent objectivity of the audience of justification.

In this way the strategy appears to be circular while also, paradoxically, reliant on consent or consensus. I would suggest that this seeming paradox is a consequence of the difficulties that arise when attempting to justify principles through consent whilst also recognising that considerations prior to consent are fundamental. Such a dilemma is the inevitable outcome of grounding justification in the acceptance of principles and conceptions that are left without independent vindication.

In Rawls's case, actual endorsement or affirmation by particular persons (specifically citizens) is assumed to offer the necessary objectivity, leading to a justificatory impasse. In contrast to Rawls's assertions, it does not appear that his account can do without a more universal strategy of justification if it hopes to demonstrate anything more than the factual acceptance of certain principles of justice in a particular context. For example, as we will see when we look at decent regimes and the international context of justice, the factual acceptance of principles of freedom and equality is insufficient as a form of justification and undermines the ability to argue for such principles in the context of a plurality of perspectives on justice.

Since consent is fundamental to Rawls's strategy of justification, agreement or consensus may define the moral significance of the content of justice. Endorsement by particular persons (conceived as "reasonable" and "objective") is inseparable from fundamental justification and, therefore, the danger is that this perspective may define the value and meaning of principles of justice. Therefore, or
correspondingly, outside the consensus, no justification for the content of justice, e.g. on the conception of what is reasonable, is available. This is because the relevant "objective" perspective is absent or unavailable. Hence such a strategy of justification must inevitably remain context-dependent; outside its particular context the grounds for viewing principles as morally significant are unavailable. This is not to say that particular contexts are not vital for embedding principles of justice, only that particular contexts cannot be fundamental when attempting to ground the justification for distinguishing who is actually owed justification.

Such a strategy has no claim to ground moral-universal duties or obligations; no one outside the consensus can possibly be bound by any moral duty or obligation. Such a consequence seems unappealing in the actual contexts in which we find ourselves, of pluralistic societies and a pluralistic international community, where agreement cannot be taken as a presupposition of reasoning about justice.

Rawls refers to those outside a consensus on particular principles of justice, such as freedom, equality and reciprocity, as "unreasonable". In this way it may be thought that a claim of moral-universal validity still holds; if all those outside the consensus can be designated as unreasonable in a morally significant sense, then the consensus might still be presumed to represent and delineate the scope of moral value or validity. Of course, such a strategy must take care to ensure that reasonable others are not also excluded, as Rawls himself commented:

We avoid excluding doctrines as unreasonable without strong grounds based on clear aspects of the reasonable itself. Otherwise our account runs the danger of being arbitrary and exclusive. (PL, 59)

And as for the relevant aspects of reasonableness:

---

80 O'Neill argues for the importance of contextualising abstract principles: "[a]n abstract deontological liberalism cannot guide action or be accessible to agents of varying Sittlichkeit unless abstract principles can be connected to determine judgments that take account of the actual identities and capacities of agents... The move from abstract principle to determinate interpretation in a given context is part and parcel of all ethical reasoning (indeed all practical reasoning)". O. O'Neill, "Ethical Reasoning and Ideological Pluralism", Ethics, Vol. 98, No. 4 (July, 1988), pp. 705-722, p. 720.
To conclude: reasonable persons see that the burdens of judgment set limits on what can be reasonably justified to others, and so they endorse some form of liberty of conscience and freedom of thought. It is unreasonable for us to use political power, should we possess it, or share it with others, to repress comprehensive views that are not unreasonable. (PL, 61)

And finally:

The only comprehensive doctrines that run afoul of public reason are those that cannot support a reasonable balance of political values. Yet given that the doctrines actually help support a reasonable balance, how could anyone complain? What would be the objection? (PL, 243-4)

Rawls is aware that without strong and legitimate grounds for designating something as unreasonable there is a serious risk that some potentially valid points of view may be arbitrarily excluded. He then defines the unreasonable in terms of the unreasonable use of political power to suppress liberty of conscience and freedom of thought. Finally, he asks: on what reasonable ground could anyone object to such political values?

Of course, the problem with Rawls’s account is not that there might be legitimate grounds on which to object to principles of liberal justice, rather the issue is that no reason has been offered to those who object to accept the conditions of the original position, that is, no reason that does not merely state that such principles are valuable because “we” recognise them to be so, that they are those “we” “would want” under conditions that could be perceived as arbitrary or relative by those who disagree. The claim that there are no reasonable grounds on which to object can hold only so long as what would constitute a reasonable objection has been justified in terms that have some objective validity both to those who agree and those who disagree. Clearly a justification based on a consensus is inaccessible to those outside such a consensus.\textsuperscript{81} Similarly to the case of decent societies, the framework for dissent (for defining unreasonableness) is framed in terms defined by the perspective of those within the consensus. In other words, suggesting that

\textsuperscript{81} This claim regarding the non-accessibility of shared premises in contexts of pluralism will be looked at in more detail when we analyze O’Neill’s account of public reason in Part 3.
there are no reasonable grounds for objection can only be a valid claim so long as you can assume an objective, i.e. non-partisan, account of reasonableness. This of course does not mean that the other must actually agree or consent to the framework, but that the framework can in principle be justified to them. As we have seen, Rawls's account of reasonableness receives fundamental justification via affirmation from those within the consensus. A strategy that implicitly relies on the actual consent or agreement of persons within a consensus cannot ground an independent perspective that seeks to make claims of unreasonableness about those outside the consensus. A sufficient justification for such designations would need to rely on something other than the shared agreement of those who seek to exclude.

It is on this point, regarding the designation “unreasonable”, that confusion may arise, because there is a slide between, or a conflation of, what can be accepted and what would be or is accepted. Whilst Rawls recognises that there are conditions that identify what can be accepted, crucially he only focuses on either what would be or what is accepted. In understanding the objection to Rawls's account we must differentiate between related but distinct claims regarding the justification of principles of justice (in this case liberal principles). Of such principles it can be claimed that they can be justified, that they cannot be justified, that they are justified and/or that they are not justified.

For this reason it may seem tempting to suppose that Rawls has provided sufficient justification, because (it is likely that) “we” (liberal citizens) intuitively accept that the principles of fairness and reciprocity are those principles that can be justified to others. Therefore, when confronted with an “other” who rejects such principles, it seems acceptable to suppose that this other may be properly described as unreasonable. While this may be true, the crucial point is that Rawls has not justified liberal principles, he has only succeeded in demonstrating that they are those principles that some of “us” do in fact accept; he does not attempt to demonstrate that they are those that can be accepted, he only attempts to demonstrate that they are those that would be accepted, based on conditions that are accepted by some. So the objection to Rawls does not state categorically that
the principles of liberalism cannot be justified to others, rather it argues that a) Rawls does not justify them, and b) that they can and must be justified. The objection can be formulated as the claim that, if we take principles of justice seriously, it is not sufficient to hold the principles of justice in coherence with (intuitively accepted principles or) convictions; if more robust and universal justification is not available then the principles themselves are undermined and demoted to the status of context-dependent, contingent “principles” potentially based on changeable cultural values.

The problem with relying on Rawls’s account of reasonableness in political liberalism to designate what is unreasonable is that this account is valid only within the consensus; outside the consensus it is lacking adequate justification and cannot function as a measure of reasonableness. His account of unreasonableness fails to deal with the most important issue, that those who object can reject the content of the account of reasonableness as not justified to them. It is not then sufficient to reject their point of view on the basis that they do not endorse those

---

82 This claim relates to principles of freedom and equality. There is a reservation regarding the principle of reciprocity, which is understood as presupposing a legal-institutional context and as therefore unable to justify such institutions. There is a more fundamental level of moral reasoning that precedes the institutionalisation of reciprocity and grounds an asymmetrical, agent-centered and obligation-based account of justice. In “Political Liberalism and Public Reason”, O’Neill notes with regard to the “condition clause” in the definition of reasonableness, that “[p]ersons are reasonable ... when ... 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”, that “[t]he conditional clause in this formulation appears, as Rawls points out (PL, 50), to identify reasonableness very closely with a commitment to a form of second-order reciprocity, or rather reciprocity about second-order principles, and to be mute where no such reciprocity can be expected. A conditional willingness to agree if there are terms that will (or would) be agreed on by all binds nobody when the condition does not obtain. Yet we would often think that certain standards and principles were reasonable even without such assurance—and others unreasonable even with that assurance.” O’Neill, “Political Liberalism and Public Reason”, p. 415. Other comments link the reciprocity condition of reasonableness to the inadequacy of a reliance on agreement: “If we did not assume that the context of reason is a closed society of democratic citizens, we might well have doubts about a conception of reasonableness that is conditional on an assurance that others will agree on the principles and standards we propose. Looked at from the perspective of outsiders, others’ agreement might be thought to reflect matters quite other than their reasonableness. Their refusal to agree might be thought to reflect either the diversities of identities or conflicts between real interests, rather than their unreasonableness. Their willingness to agree might be thought an indicator of the rawness of domination or of compliant subservience rather than a condition of their reasonableness. Some cases of reciprocity may not be reasonable.” O’Neill, “Political Liberalism and Public Reason”, p. 423.
principles. Without demonstrating the universal (i.e. non context-dependent) validity and relevance of the principles of justice, such as freedom and equality, the other can simply respond that one’s position is not endorsed from their ethical perspective, and is therefore not justified from their perspective. Only a justificatory strategy and perspective that is independent of the consensus can show those outside the consensus are unreasonable based on a universally, and hence non-partisan, valid critical standard, i.e. one not dependent on a particular perspective.

In order to seek a more appropriate method of justification, that is not reliant on a consensus, it must be recognised that something other than consent is fundamental to justification. Consent cannot demonstrate obligatory action (by way of justifying unconditional principles), it can only demonstrate the permissibility of action, and even then it can only do so once a prior stage of justification has occurred. While hypothetical consent goes further than actual consent in attempting to describe the conditions that legitimate consent, it attempts to justify the conditions via a strategy that relies on consent, and therefore fails to justify the relevant conditions. If alternative accounts, such as O'Neill’s Kantian account, that focus on the possibility of consent can succeed, they can explain why certain conditions are necessary, as opposed to only factually accepted, and in doing so can justify what Rawls sets out to justify, but in the end presupposes.

2.2. A critique of The Law of Peoples in light of the problematic status of Rawls’s account of justification

Having critiqued the justificatory foundations of Rawls’s account of justice, and concluded that Rawls’s strategy relies on actual agreement and as such is grounded in an insufficient form of justification, I turn to The Law of Peoples to look at the implications this consent-dependent justification has on his account of international justice. This section will begin with a brief critique of the claim that decent regimes, i.e. regimes that reject principles of freedom and equality, endorse
fairness and equality at the international level (2.2.1.). I will argue that Rawls’s reasoning on the subject moves in the wrong direction of causality, and as such avoids the question of why such regimes would, on any substantive ethical level, endorse a principle at the international level while rejecting it at home. I highlight this not to argue that such an endorsement is impossible, but to highlight how Rawls’s account of decent regimes contains some apparent inconsistencies that bear on the likelihood of establishing a just Society of Peoples in this way.

After addressing Rawls’s proposals on the endorsement of principles of equality, I turn to address the simultaneous homogeneity and plurality of perspectives on ethics and justice within decent regimes, which may turn out to be a more serious inconsistency in the description (2.2.2.). With regard to homogeneity, I argue that Rawls intends for decent regimes to be understood as societies wherein the vast majority of members share a common perspective on justice, in their collective affirmation of the “common good idea of justice”. I suggest that this is not a valid assumption to be made regarding decent regimes, on the basis that they are also depicted as containing a multiplicity of ethical perspectives or comprehensive doctrines. Whereas in political liberalism reasonable plurality can (it is argued by Rawls) be neutralised in the public sphere, such a homogenisation is not available in regimes that reject the principle of reciprocity.

In the place of the principle of reciprocity, Rawls asserts a robust procedure of consultation. While the goal of urging members to engage with each other is to be commended, I question the effectiveness of this procedure on the basis that it takes place within the framework of the common good idea of justice. I argue that because it presupposes this framework, it cannot legitimise it. Based on O’Neill’s understanding of the conditions for consent as a form of agency, actual agreement to the shared idea of justice in decent regimes cannot ground legitimacy, and therefore the notion of decent regimes as a people is undermined. This raises significant problems for Rawls’s thesis, as it is the notion of decent regimes as a people that allows him to shield them from criticism from a liberal perspective, through claims of “self-respect” and “honor”.
Practical Reasoning and Transnational Justice

From our critique of the role of consent, I argue that it is now apparent that Rawls takes evidence of loyalty or affirmation of a nonliberal conception of justice as (at least partially) legitimate because on his account what persons actually agree to is fundamentally relevant to justification (2.2.3.). Rawls has not demonstrated the moral justifiability of his account outside a consensus on its content; there is therefore worryingly little ground on his account to critique others’ agreements as less legitimate, at least from their perspective. The overlapping consensus on political justice is only “objective” from within the liberal perspective, i.e. from within the consensus, and there is therefore no critical ground on which to critique others’ agreements on conceptions of justice as unjust. Rather, any critique of alternative (and possibly unjust) political systems can be rejected by those others as merely the exportation of liberal values, or a form of cultural imperialism.

This, I argue, is the ultimate impetus behind Rawls’s particular account of decent regimes (2.2.4.). The “objective perspective” that justifies principles of justice is only available within the consensus. Rawls therefore faces a dilemma between arguing for such principles beyond the limits of the consensus or else respecting the (possibly equally legitimate) agreements of others outside the consensus; at the international level it appears he must choose between liberal principles and “respect” for other consensus-based accounts of justice. I suggest that Rawls’s fundamental strategy is therefore to describe nonliberal regimes that, while not meeting liberal standards, are not sufficiently unjust to warrant criticism from a liberal perspective (such criticism, recall, cannot claim objectivity). However, Rawls fails in this regard: it cannot be presumed that decent regimes as described are grounded in a legitimate consensus on a nonliberal conception of justice. I argue, based on my analysis of O’Neill’s critique of consent-based justification, that freedom and equality are necessary conditions of consent, and in the absence of these conditions no apparent agreement on a conception of justice can be accepted as legitimate. I conclude, in agreement with Forst, that, in reality, regimes that reject basic principles of freedom and equality would stand in need of critique, and therefore it is necessary to construct a properly objective account of the principles of liberal justice, in the sense of one that is capable of justifying itself to those outside the consensus.
Decent regimes are not sufficiently legitimate to warrant the bracketing of criticism of their political institutions. On this basis, I examine whether they can function as legitimate agents in the international original position (2.2.5.). I argue that for a regime to be seen as a collective unit for reasoning about international principles of justice, the perspective of individual members must be viewed as interchangeable in the public sphere, i.e. in terms of reasoning about public political justice. I argue that while a case may be made for this within liberal regimes (though I will critique this claim when we look at Kantian public reason in Part 3), such an assumption cannot be made within decent nonliberal regimes. Rather, the appropriate assumption is that there exist a plurality of (potentially reasonable) perspectives on justice. Such perspectives cannot be assumed to have been reconciled because the regime does not meet the conditions for legitimating agreement on an apparently shared conception of justice. With regard to decent regimes, I therefore conclude that such states cannot be taken as legitimate agents for reasoning about justice at the international level.

2.2.1. “Nonliberal” conceptions of justice and the endorsement of equality at the international level

Rawls first comments on the features of decent societies that leads them to endorse principles of international equality when he sets out the criteria for a decent regime to be a member of a Society of Peoples. The first criterion states that

First, the society does not have aggressive aims, and it recognizes that it must gain its legitimate ends through diplomacy and trade and other ways of peace. Although its religious or other underlying doctrine is assumed to be comprehensive and to have influence on the structure of government and its social policy, the society respects the political and social order of other societies. If it does seek wider influence, it does so in ways compatible with the independence of other societies, including their religious and civil liberties. This feature of the society’s comprehensive doctrine supports the institutional basis of its peaceful conduct and distinguishes it from the leading European states during the religious wars of the sixteenth and seventeenth centuries. (LP, 64. Emphasis mine)
The crucial statements here are that such a regime "does not have aggressive aims", "respects the political and social order of other societies", and that this is a "feature of the society's comprehensive doctrine". To begin, although the list of desirable features sounds compelling, this description fails to specify the reasons and motivations why such a society has non-aggressive aims and respects the integrity and sovereignty of other regimes. Such regimes appear to be just those who happen to have such aims; there is little to connect the form of reasoning that guides their comprehensive doctrine to their policy of non-aggression and their endorsement of international principles of equality and fairness. Indeed, there seems little reason for thinking that a regime that rejected a principle of equality between persons could endorse such a principle between other types of agent. The only explanation Rawls provides for why decent peoples are capable of making such a distinction (between forms of reasoning) is that they have non-aggressive aims toward other peoples and respect their social order. So again the question arises as to why they have the capability and the motivation to have non-aggressive aims toward other societies.

I suggest that this peculiar conflation of non-aggressive aims with the endorsement of principles of equality is the result of an inversion of causal reasoning in Rawls's thinking. I suggest that Rawls inverts the direction of causality between non-aggressive aims and forms of reasoning.

In other words, Rawls implies that because decent regimes have (for some unspecified reasons) non-aggressive aims, they ought therefore be capable of viewing other states as equal in the international sphere. However, non-aggression, as a foreign policy attitude, is the outcome of a process of reasoning and not its cause or origin. A regime that endorses principles of equality, based on a particular account of reason, will most likely pursue peaceful foreign policies. This idea may not however be inverted in order to claim that non-aggressive aims imply a capacity to endorse principles of equality, especially where such principles are rejected internally. If endorsing principles of equality implies non-aggression, then it must still be explained why principles of equality are endorsed. I suggest that
this inversion of causality allows Rawls to avoid the question of why a regime whose domestic account of reason and justice rejected equality between persons would still endorse such a principle between peoples.\textsuperscript{83}

This form of reasoning leaves unvindicated or at least unexplained the premise that decent regimes might or could have non-aggressive aims. This seems further supported by Rawls's comment that peaceful aims are a feature of the comprehensive doctrine/s of such regimes. Decent regimes appear to be those whose doctrines happen to lead to policies of either non-aggression or respect for other societies while, paradoxically, not endorsing fairness or equality toward persons.

Here I draw an analogy to Rawls's thinking on comprehensive doctrines and public reasoning in \textit{Political Liberalism}. In \textit{Political Liberalism}, comprehensive doctrines within a liberal society overlapped on certain core values of political justice that were grounded in fairness and reciprocity. As argued in the previous section, Rawls does not specify with clarity whether such an overlap was the result of a common and shared capacity for reasoning about justice or whether it was the result of a chance coincidence of values all originating from diverse systems of reasoning. While on the one hand he asserted that diverse ethical systems of thought were impenetrable, on the other hand he argued that all such doctrines could understand and endorse the principles of practical reason; citizens were portrayed as simultaneously reasonable in their ability to translate their comprehensive doctrine into a political conception with overlaps to others, yet unreasonable, in that Rawls suggests they cannot engage constructively with each other on ethical questions. Rather a common ground of shared premises must be sought because one could not argue from premises the other did not \textit{already} have access to or endorse in their comprehensive doctrine.

\begin{footnote}
\textsuperscript{83} That is excluding a reference to self-interest. While he does claim that decent regimes have a self-interested concern to agree to principles of equality, he also implies, by describing them as non-aggressive, that some aspect of their account of reason and justice leads them to an (at least partially) ethically based endorsement of equality between peoples. It is this aspect that appears to be inconsistent with his description of the domestic account of justice of decent regimes.
\end{footnote}
In the case of decent regimes, it is again unclear whether they endorse principles of equality and fairness at the international level because they share with liberal regimes, at least on some level, a common capacity for reasoning about justice, or whether it is merely pure luck that there exist other forms of reasoning that are radically distinct from political liberalism, yet happen to be capable of endorsing a subset of the principles of practical reason set out by Rawls. As before, it seems that if there exists a shared capacity for reasoning that allows decent regimes to endorse principles of equality then such regimes must be capable of at least comprehending an application of that form of reasoning to their internal domestic institutions, or at least capable of comprehending criticism of their political institutions based on a shared account of practical reason. If, on the other hand, distinct forms of reasoning are impenetrable then there is no reason for expecting such regimes to endorse principles of equality and fairness at the international level. It seems that, if we can seek any form of agreement at the international level, then disparate forms of reasoning must share some common ground and we may at least attempt to communicate with and critique alternative systems of thought.

The notion of critique, however, is problematic for Rawls; free communication leads to irreducible disagreements. For this and other reasons Rawls’s account of international justice circumvents communication by seeking out alternative systems of thought that are already at least partially acceptable from the perspective of liberal peoples. However, what the above critique of the endorsement of equality has shown is that there are identifiable inconsistencies in Rawls’s account of “acceptable” “nonliberal” political institutions. Further inconsistencies, to which we know turn, relate to the simultaneous homogeneity and plurality of perspectives on justice within decent regimes.

2.2.2. The neutralisation of plurality in the public sphere of liberal and decent societies

With regard to the agreement on a shared conception of justice in decent and liberal regimes, Rawls invokes the concept of a “people” to suggest a level of
homogeneity that allows him to treat collectivities of persons as agents in a similar sense to individual persons. By “homogenise” I refer to attempts to reconcile a diverse and possibly disparate range of perspectives on justice and bring them into a shared realm of acceptance. For now, I address only whether the attempts implied by Rawls are capable of neutralising ethical plurality in a manner that could ground the assumption of a shared perspective on justice in the public sphere.⁸⁴ There are several comments Rawls makes which emphasise how “peoples” are to be understood as homogenous entities, in particular his discussion of the “self-respect” of such entities.⁸⁵

He warns liberal peoples that intolerance will “wound the self-respect of decent nonliberal peoples” (LP, 61), and elsewhere claims that peoples have a “definite moral nature” and are capable of feeling “pride” and “honor” (LP, 62). The only practical inference from these claims is that they apply to the entirety of a people and not merely to the ruling class or those positioned at the apex of a hierarchical structure. If a substantial portion of the population did not experience such feelings there would be little reason to attribute them to a “people”. Also Rawls refers to decent peoples as “sincerely affirming a nonliberal idea of justice” (LP, 70. Emphasis mine). And with regard to the suggestion that a diversity of affinities or associations may exist, Rawls suggests that for now we may assume a level of commonality:

---

⁸⁴ I do not address the question of whether the best way of dealing with ethical diversity is to seek a homogenised perspective in the public sphere in this way; whether a “shared” public realm must be restricted by the shared perspective on justice that Rawls envisions. In Part 3 I address the limitations of Rawls’s account of public reason for grounding a shared account of reasoning in contexts of ethical plurality.

⁸⁵ Here Rawls applies the criterion of self-respect to states. In A Theory of Justice he discusses self-respect as an extremely important criterion and suggests it may be “the most important primary good”: “On several occasions I have mentioned that perhaps the most important primary good is that of self-respect. We must make sure that the conception of goodness as rationality explains why this should be so. We may define self-respect (or self-esteem) as having two aspects. First of all, we noted earlier (§29], it includes a person’s sense of his own value, his secure conviction that his conception of his good, his plan of life, is worth carrying out. And second, self-respect implies a confidence in one’s ability, so far as it is within one’s own power, to fulfil one’s own intentions. When we feel that our plans are of little value, we cannot pursue them with pleasure or take delight in their execution. Nor plagued by failure and self-doubt can we continue in our endeavours. It is clear then why self-respect is a primary good. Without it nothing may seem worth doing, or if some things have value for us, we lack the will to strive for them. All desire and activity becomes empty and vain, and we sink into apathy and cynicism. Therefore the parties in the original position would wish to avoid at almost any cost the social conditions that undermine self-respect. The fact that justice as fairness gives more support to self-esteem than other principles is a strong reason for them to adopt it” (TJ, 440).
Practical Reasoning and Transnational Justice

Notwithstanding, the Law of Peoples starts with the need for common sympathies, no matter what their source may be. My hope is that, if we begin in this simplified way, we can work out political principles that will, in due course, enable us to deal with more difficult cases where all citizens are not united by a common language and shared historical memories. One thought that encourages this way of thinking is that within a reasonably just liberal (or decent) polity it is possible, I believe, to satisfy the reasonable cultural interests and needs of groups with diverse ethnic and national backgrounds. We proceed on the assumption that the political principles for a reasonably just constitutional regime allow us to deal with a variety of cases, if not all. (LP, 24-5. Emphases mine)

Peoples on this account are capable of feeling certain emotions, such as pride and shame, they exhibit a measure of “common sympathies” and, perhaps most crucially, they “sincerely affirm” a particular conception of justice. On the basis of these attributes Rawls designates peoples as agents capable of reasoning as a coherent and unified entity. A “people” as agent is an amalgam of the individual perspectives that constitute the collective agent. This of course does not imply that the individual members agree on all matters, but does imply that at least on larger political matters, and on international law, they can be taken as presenting one coherent perspective.

I suggest, however, that this picture is not consistent with the necessary implications of other features of the description of decent regimes. Firstly, decent societies “fail to treat persons who possess all the powers of reason, intellect and moral feeling as truly free and equal”, although they do allow them a “substantive political role” and a “right of dissent” (LP, 61). This is because a decent hierarchical society’s conception of the person, as implied by the second criterion, does not require acceptance of the liberal idea that persons are citizens first and have equal basic rights as equal citizens. Rather it views persons as responsible and cooperating members of their respective groups. Hence, persons can recognize, understand, and act in accordance with their moral duties and obligations as members of these groups. (LP, 66)

Though they are not equal, they are seen as “decent and rational”, “as capable of moral learning in their society” (LP, 71) and as playing “a certain role in the overall scheme of cooperation” (LP, 72). They must also be allowed “a sufficient measure
of freedom of conscience and freedom of religion, even if these freedoms are not as extensive and equal for all members... as they are in liberal societies.” (LP, 72)

Therefore, based on a conception of the person that does not begin from the perspective of Political Liberalism by taking citizens as the bearers of equal political rights and duties, decent societies are structured hierarchically with some members being considered less equal than others, and furthermore, this positioning along the hierarchical scale is based on membership in distinct groups who endorse distinct comprehensive doctrines. This is evident from his description of the fictitious regime of “Kazanistan”, a country where Islam is the religion of the state. In this state Islam has priority, although other religions have “most civic rights” (LP, 76). Despite the multiplicity of comprehensive doctrines, most members of the regime accept their society and their role within it:

These minorities have been loyal subjects of society, and they are not subjected to arbitrary discrimination, or treated as inferior by Muslims in public and social relations... The Muslim rulers have long held the view that all members of society naturally want to be loyal members of the country into which they are born; and that, unless they are unfairly treated and discriminated against, they will remain so. Following this idea has proved highly successful. Kazanistan’s non-Muslim members and its minorities have remained loyal and supported the government in times of danger. (LP, 76-7)

The question of whether such loyalty can ground a legitimate affirmation of a conception of justice is one to which I will return. For now what is relevant is the hierarchical nature of the regime, and the distinct identities of the members within the regime based on their distinct comprehensive doctrines. Whereas, from Political Liberalism, in liberal societies the distinct private identities of citizens are neutralised in the public sphere by their shared public identity as equal political citizens employing a common public reason grounded in an overlapping core of shared values, no such equivalence exists in the public sphere of decent regimes. Rather, their form of public reasoning draws on comprehensive doctrines to distinguish individuals in the public sphere, or more specifically to distinguish members of groups, for even minorities cannot be a minority of one.86 In Political

---

86 Among Rawls’s guidelines for a consultation hierarchy is the requirement that “each member of a people must belong to a group”. (LP, 77)
Practical Reasoning and Transnational Justice

*Liberalism* the common public reasoning is formed via the overlapping consensus, which gives a common citizens' point of view and serves as the basis for grounding claims of objectivity in public reasoning. Given the distinct public identities of the members of decent regimes, it appears that no such common point of view can be claimed, and therefore that no discussion of the reasoning of decent regimes can claim to reason from the point of view of a “member” of a decent regime. Rather, such members are distinct agents in the public realm of their society whose perspectives have not been neutralised as in the liberal society through seeking out shared premises. The point of view of one member cannot be taken as the point of view of another, and this distinction is compounded given that the disparate perspectives are situated along a hierarchical scale.

Rawls seems therefore to depict decent regimes as simultaneously homogenous and pluralistic. They are homogenous in their shared feeling and affinity that make them capable of feeling “wounded” or proud depending on their treatment by other regimes, yet also constituted by a plurality of comprehensive doctrines which are not situated equally and who therefore must hold distinct perspectives in relation to the public sphere, and possibly also therefore toward the international public sphere, thereby seemingly contradicting any claim to share any feeling in relation to other regimes. In *Political Liberalism*, liberal societies manage their pluralism by homogenising public reason through an overlapping consensus grounded in a principle of reciprocity between comprehensive doctrines that are assumed to be both disparate and combinable. In the case of the decent society, Rawls must seek a different way to homogenise the perspectives of their members.

2.2.3. The idea of a consultation procedure to legitimise the affirmation of a nonliberal account of justice, and its problems

As we saw above, Rawls describes decent “peoples” as “sincerely affirming a nonliberal idea of justice” (LP, 70). In the example of Kazanistan this is explicated further; as long as members are not “unfairly treated”, they will “have remained loyal and supported the government in times of danger” (LP, 76-7). The basis for
the non-rejection or affirmation of the nonliberal conception of justice by the members of decent regimes, in particular those members who are viewed as less equal than others, is their loyalty to the regime, based on their perception of their treatment as "fair". On this basis can they be conceived of as a "people" with sufficient commonality of perspective to be treated as a distinct reasoning agent in the Society of Peoples.

However it must be asked precisely what constitutes fairness in nonliberal regimes. While the conception of fairness deployed by Rawls throughout most of his work relies on the criterion of reciprocity to ground its objectivity, in decent regimes, the connection to reciprocity has been severed. In its place it seems that fair treatment is now related to the procedures of consultation that allow objection and dissent a role in the political process, i.e. that persons can be thought of as fairly treated when they have the opportunity to voice concerns and objections relating to their political system. In democratic societies, consent for the regime is considered legitimate because it is produced via democratic procedures of representation. Free and uncoerced agreement implies the validity of the result, because those who object have the opportunity to reject or renegotiate the proposal. Where persons have or had the capacity to refuse, their consent can be considered legitimate.

The procedure of consultation Rawls depicts is intended to offer an account of alternative procedures for dissent sufficient to at least partially legitimise consent or agreement. In decent regimes those who object have the opportunity to express their dissent in the consultation process with the realistic expectation that their dissent will be taken seriously by political leaders, judges and government officials. If this procedure is sufficient to take dissent seriously, this implies that the outcomes of procedures of consultation can be considered inclusive, free and uncoerced, and are therefore to a certain extent legitimate. Where dissent is an available option, the collective affirmation of the regime cannot be considered coerced and therefore legitimates the postulation of a shared conception of nonliberal justice; the consultation procedure is vital to the account of decent regimes; it defines the perception of "fair" treatment that grounds the description
of members as loyal to the regime. Hence Rawls's effort to describe a system where equal representation is not legally guaranteed yet dissent is still taken seriously.

However there are certain considerations that call into question the suggestion that such procedures are sufficient to legitimise any form of "affirmation" of the regime. These problems relate to the presumption of a common perspective on justice in the public sphere of decent regimes.

According to Rawls, a decent society's system of law is guided by a "common good idea of justice", which he distinguishes from a "common aim of a people" (LP, 71). The important difference between the two conceptions is the "procedure of consultation". The distinction he draws seems to reflect the assertion in Rawls's earlier works that the rational must be circumscribed by the reasonable, and his criticism of Utilitarianism, in particular that the pursuit of the common or aggregate good cannot be at the expense of individuals. He writes that

the pursuit of the common aim is to be encouraged, but is not to be maximized in and of itself, but rather maximized consistent with the restrictions specified by honouring the steps in the consultation procedure, which provides the institutional basis for protecting the rights and duties of the members of the people. (LP, 71)

The significance of the distinction then is that the procedure of consultation regulates the pursuit of any common aims a people might have by attempting to take each group's perspective into consideration. Although the procedure does not consider members to be "separate individuals deserving equal representation" (LP, 71), they are seen as "rational" and as "capable of moral learning" and as "able to recognize when their moral duties and obligations accord with the people's common good idea of justice" (LP, 71). The procedure "allows an opportunity for different voices to be heard", not in a democratic sense but "appropriately in view of the religious and philosophical values of the society as expressed in its idea of the common good" (LP, 72. Emphasis mine). Persons, as members of groups, "have the right at some point in the procedure of consultation... to express political dissent" (LP, 72). And the right of dissent does not stop at expression, rather "[j]udges and
other officials must be willing to address objections", and they do so as a result of
their “sincere belief” that “the justice of the legal system must include respect for
the possibility of dissent” (LP, 72). Finally, “dissenters are not required to accept
the answer given to them” and may “renew their dissent” (LP, 72), though the
power to ultimately refuse the demands of dissenters and proceed with particular
political action presumably lies with the judges and political officials.

Political dissent, or more generally political debate, requires at least an account of
reasoning, or in Rawls’s work a framework for public reasoning. Rawls finishes his
account of dissent in decent regimes with the claim that “[p]olitical dissent
expresses a form of public protest and is permissible provided it stays within the
basic framework of the common good idea of justice” (LP, 72. Emphasis mine).
Therefore the framework for political discussion is the common good idea of
justice. This framework is a shared framework and can therefore function as a
shared basis for grounding dissent and objection, much as in the case of public
reason in liberal societies, though to a lesser extent. If it were an imposed
framework, it is difficult to see how or why objection or dissent could be framed
by it. That it is intended as a collectively endorsed framework is reinforced by
Rawls’s claim that a people affirms a nonliberal idea of justice, implying a people as
a collective agent and not merely referring to the dominant group.

However, it follows from what I have argued so far that the common good idea of
justice cannot be taken as a shared framework for reasoning about justice in
decent societies. This is not directly because of the more obvious criticism that
such a framework denies equal standing to all members, but because, as I noted
above, such a framework expresses “the religious and philosophical values of the
society”. From our analysis of the evidence of a level of pluralism in decent
regimes (that there are minority and majority religions or doctrines situated
according to hierarchical groupings), it is clear that there exists a diversity of
religious and/or philosophical value systems in Rawls’s decent society, and that one
set of particular religious or philosophical values is dominant. Given that the
political system is defined by the values of the dominant set, this dominant set of
values are those that have ultimate control over the formulation of the common good

133
idea of justice. For example, in Kazanistan, Islam is the dominant religion and can specify the status of other religions in relation to its position. This interpretation is further reinforced when we recall that the affirmation that grounds the “shared” nature of the conception is grounded in “loyalty toward the regime”, and not in any form of political self-determination on the part of the members of decent regimes.

Therefore, the values that shape or frame the common good idea of justice do not represent the diversity of values that exist in the society (according to Rawls only a neutral political conception can achieve this reconciliation of comprehensive and general perspective), rather they are those of a particular group. Therefore what is called the common good idea of justice is in danger of being a framework imposed by a particular perspective upon other subordinate perspectives.

A response may be that Rawls fully acknowledges that the different groups are not situated equally, and that one doctrine dominates, however this response is misleading. What I am attempting to highlight is the inconsistency in this description: Rawls admits that minorities are not treated as equal but also assumes and suggests that the common good idea of justice is to some degree a shared framework and can therefore serve as the framework for expressing dissent. I argue that these two assumptions are incompatible. Again this is because Rawls implicitly depicts decent regimes as both homogenous, in the sense of sharing a public conception of justice (as a “people” whose members are “loyal” to their regime), and pluralistic, while failing to neutralise the plurality of perspectives as he claims to do with the political liberal conception of justice, thereby failing to ground any shared framework within which legitimate dissent could be expressed.87

87 Though again this does not imply that his strategy of neutralisation of perspectives in the public sphere of liberal democracies is successful or desirable. His account of public reason in liberal democracies specifies in advance the framework, or content of public reason, and cannot be challenged during the process of political debate. This will be explored in greater detail when we turn to O’Neill’s alternative account of public reason. Also, given that individuals must belong to particular groups or traditions, it must also be asked by what means may individuals critique their group? Presumably the group has its own internal procedures for managing dissent, which brings to mind common objections to communitarianism relating to the potentially power-laden nature of such procedures. It seems that members of decent regimes are at risk of being doubly caught in imposed and unchallengeable frameworks for debate and dissent, both at the group and state level.
What this implies for legitimate dissent is that dissenters (in particular those from subordinate groups) are forced to articulate dissent within a framework imposed by an alternative set of religious or philosophical values, and more crucially cannot critique this framework because there is no account of reason available upon which to mount such an objection. Indeed it may well suit the dominant doctrine that those with other value systems or comprehensive doctrines must find a way to express criticism of the regime within a structure defined by the majority or ruling doctrine. This is not something Rawls would consider appropriate for a “decent” regime, however I suggest that this is something Rawls overlooks by implicitly conflating a common good idea of justice with a legitimately shared conception of justice. It is this common good idea of justice that allows Rawls to characterise certain regimes as decent, and I suggest that this is because of the implied assertion that such a conception of justice is to some degree a shared conception. And crucially it is this assertion that allows him to mitigate the unjust effects of the denial of full equality to some members of decent regimes.

2.2.4. Between justice and respect: Rawls’s paradoxical account of decency as the outcome of his context-dependent strategy of justification

A question that arises from our critique of decency The Law of Peoples is why apparent loyalty towards a particular regime should constitute grounds to consider that regime sufficiently legitimate to warrant respect, in the sense of withholding criticism from a liberal perspective. What I suggest is that, if actual consensus or consent forms a crucial and fundamental aspect of justification in Rawls’s thesis, he must then treat actual consent to or endorsement of other sets of principles of justice as (at the very least partially) legitimate. If consent can justify, then the fact of apparent consent to other political regimes must be taken seriously. Rawls’s understanding of legitimacy can be linked to his inability to take the conditions for consent, as opposed to consent itself, as fundamental to justification.
What must be emphasised regarding this point is that, on Rawls’s account, consent, in this case to potentially unjust regimes, must be taken seriously despite the fact that the conditions that legitimate the consensus in political liberalism have not been met. While it may seem that on Rawls’s account actual consent to nonliberal systems does not confer legitimacy because agreements of this kind are detached from conditions of reciprocity (that are a part of the conception of reasonableness), in fact, as we have seen, reciprocity and the conception of reasonableness from which it is derived are only justified by reference to a balancing of its implications (the principles of justice as fairness) with an acceptance or affirmation of its validity. Therefore, as we noted, the conditions that might legitimate consent as morally justified are themselves justified via a reliance on agreement, making consent, and not its legitimating conditions, fundamental. If the conditions are not fundamental, i.e. if the conditions cannot be justified independently from and without recourse to consent, Rawls cannot claim that their absence nullifies claims that consent or affirmation itself grounds claims of legitimacy. Rawls’s willingness to take actual agreements in decent regimes as sufficient for legitimacy despite their separation from just conditions derives from an inability to justify his own conceptions without reference to agreement.

A corollary of this point is that on Rawls’s account there are actually troublingly few grounds on which to treat “nonliberal” institutions as either more or less legitimate than liberal institutions. Outside the consensus, conditions of reciprocity or equality receive no justification; therefore, there are no objective or independent grounds on which to designate the “other” who rejects such principles as unreasonable. I would suggest that this interpretation fits with Rawls’s discussion of the need for toleration. His comments on toleration highlight his recognition that, as a result of its justificatory strategies, political liberalism cannot transpose its principles onto other cultures without risking a form of cultural imperialism. Of the universal moral relevance of liberal principles in the

---

88 Rawls denies that his account of international justice is a form of cultural imperialism. In concluding The Law of Peoples Rawls comments: “To the objection that to proceed thus is ethnocentric or merely western, the reply is: no, not necessarily. Whether it is so turns on the content of the Law of Peoples that liberal societies embrace. The objectivity of that law surely depends not on its time, place, or culture of origin, but on whether it satisfies the criterion of reciprocity and belongs to the public reason of the Society of liberal and decent Peoples” (LP, 121).
international context Rawls comments that “[i]f all societies were required to be liberal, then the idea of political liberalism would fail to express due toleration for other acceptable ways (if such there are, as I assume) of ordering society” (LP, 59).

He then compares the toleration of “decent” regimes to the toleration in liberal societies of reasonable comprehensive doctrines89 and suggests that to attempt to export liberal principles would be to deny decent regimes a “due measure of respect” (LP, 61). It is intolerant then to attempt to argue for liberal principles in a nonliberal context because there are other potentially acceptable ways of ordering societies.

Forst questions the very structure of Rawls’s approach to international justice and human rights: he suggests that taking as one’s starting point the perspective of a liberal society misunderstands the challenge that human rights pose to peoples that deny basis conditions of freedom and equality:

And even if Rawls in The Law of Peoples was not guilty of locating the justification of human rights in a presumably existing or possible universal consensus, he was willing to restrict the list of human rights so that certain important rights, such as equal liberties for persons of different faiths or a right to equal political participation, were not included. One reason for this is the assumed connection between human rights and intervention just criticized, and another is the aim to respect nonliberal but “decent” peoples as worthy of being agents of justification when it comes to a common law of peoples (and to avoiding Western ethnocentrism). But the question of whether “decent hierarchical peoples” can or should be expected to conform to a “liberal” conception of human rights which is foreign to their cultural self-understanding, if asked from the perspective of the “ideals and principles of the foreign policy of a reasonably just liberal people, is misguided. For the essential question from a perspective that puts human rights first would be whether such peoples—or their governments—had legitimate reasons to deny their members equal liberties or the claim to political participation. This is what it means to say that we need to take “their” point of

However, as we have seen, the ultimate justificatory strategy for the principle of reciprocity (i.e. for the account of practical reason) relies on the actual agreement of liberal persons, and not on the conditions for legitimating agreement. It is not therefore possible on Rawls’s account to appeal to the principle of reciprocity to ground a claim of universal, or at least not Eurocentric, validity. 89 “We recognize that a liberal society is to respect its citizens’ comprehensive doctrines—religious, philosophical, and moral—provided that these doctrines are pursued in ways compatible with a reasonable political conception of justice and its public reason. Similarly, we say that provided a nonliberal society’s basic institutions meet certain specified conditions of political right and justice and lead its people to honor a reasonable and just law for a Society of Peoples, a liberal people is to tolerate and accept that society”. (LP, 59-60)
It must be noted that the objection I have made to Rawls's account does not suggest that there are not other acceptable ways, distinct from liberal political institutions, of ordering societies, but that his assertion of this fact is grounded in an inability to justify liberal principles beyond the context of the overlapping consensus, rather than in the more convincing assertion that liberalism may be one articulation of a more fundamental and pre-institutional account of practical reason and justice that can justify principles of freedom and equality. This is significant because there may by legitimate cases for arguing that the exportation of a liberal culture (in Rawls's sense) is indeed a form of cultural imperialism, but in these cases it would be because the other culture in question could be shown to embody the principles of equality and freedom necessary to legitimise a regime.

---

90 Rainer Forst, "The Justification of Human Rights and the Basic Right to Justification: A Reflexive Approach", Ethics, Vol. 120, No. 4. (2010), pp. 711-740, p. 728. Forst continues that "Rawls presupposes that a 'decent' society is characterized by a 'common good conception of justice' and by a 'decent consultation hierarchy', and thus we are to believe that there are no further claims to human rights raised since there is a high degree of internal acceptance in that society. However, if disunity and conflict were to appear in such a society and it 'cracked,' so to speak, and some members raised the claim to more demanding rights as human rights which they could justify reciprocally by attacking certain social and political privileges, would the internal authorities then have good reasons to deny these claims, and would outsiders have good reasons to say that the claims raised are not really human rights claims? I do not think so." Forst, "The Justification of Human Rights", p. 728. Forst's comments highlight an important question regarding the stability and justice of a regime that is dependent on the acceptance of institutions that are incompatible with members' self-determination. It is also interesting that it premises its objection on the possibility that disagreement may arise regarding the conception of justice, in which case the injustice of the denial of freedom and equality would be brought to the fore. Such a position could be contrasted with that of O'Neill, who argues that actual acceptance or rejection, while playing some role, is not fundamental to the justice or legitimacy of the institutions. I suspect that such difference might be trivial in practice as Forst is critical of political structures that deny freedom and equality, but it is worth noting the different weights they give to whether or not members are actually actively rejecting the political structures. On the one hand, it puts a question to O'Neill's account regarding the role of actual agreement, which from an agent-centred perspective such as hers must be taken seriously. While she does not deny its significance, the tension regarding the exact status of actual agreement with regard to conditions of possible agreement must be explored. On the other hand, it also emphasises her unique emphasis on possible consent over actual or hypothetical consent. It may appear that accounts such as Forst's, which are much closer to O'Neill's and argue for the necessity of certain principles over against their preferability, may still not sufficiently emphasise that what is actually agreed is not fundamental to justice and is not legitimate unless the underlying capacities for agency that make agreement possible are taken seriously.

91 As Forst also notes: "Human rights do not prescribe a concrete specification of the arrangements of a society. They provide a language that can be spoken in many tongues, but it is the language of emancipation. Forst, "The Justification of Human Rights", p. 729. The political philosopher and ethicist Martha Nussbaum also criticises Rawls on the basis that he takes his fundamental liberal conceptions to be "parochially Western". She argues that what is actually parochial is the assumption that such conceptions are intrinsically Western, and that it is possible
For Forst, the justification of a human right such as democracy is universal, and it misunderstands the fundamental nature of democracy to view it in terms of a conformity to particular Western political values:

[o]ne cannot limit the right to democracy by appealing to the principle of collective self-determination, for that is a recursive principle, with a built-in dynamic of justification that favors those who criticize exclusions and asymmetries. The right to democracy, as I conclude, is an undeniable right to full membership in a society, but it need not be claimed in a "liberal" sense if "liberal" means conformity to current social orders in the West. Whether the members of a society interpret and use that right in such a way that it realizes a form of liberal or egalitarian democracy is up to them (as long as these decisions are not made under pressure and indoctrination) but, given the nature of human rights as protecting and expressing the right to codetermine one's polity in an autonomous manner, there is no reason to doubt that there is a human right to democracy.

The limitation of democratic rights of participation can be understood as inconsistent with claims of a right to "collective self-determination"; such a right must imply the "right to co-determine one's polity", hence it is not possible to separate them. On respecting decent regimes, Rawls comments that

[w]ith confidence in the ideals of constitutional liberal democratic thought, [the Law of Peoples] respects decent peoples by allowing them to find their own way to honor these ideals. (LP, 122)

This seems a curious statement if we recall that decent regimes, by rejecting principles of freedom and equality, reject the basic conditions that underpin liberal value systems. As Forst notes, acknowledging that a right to equal freedom does not imply conformity to a particular set of political institutions is not the same as the claim that there are acceptable ways of ordering society that reject such basic rights, or that reject what O'Neill might refer to as conditions of free agency that make participation possible.


I would suggest that the above claim belies Rawls's reliance on what persons actually agree on to ground legitimacy and justification. Rawls's implicit assumption is that it is the fact of agreement or affirmation of an account of justice, rather than "partially" meeting conditions of justice such as freedom and equality, that makes regimes comparable to liberal regimes in the realisation of justice. Agreement, rather than the conditions of agreement, is fundamental to his concepts of justice and justification.

I suggest (bearing in mind the above) that Rawls's claim that refusing to affirm the acceptability of decent regimes is disrespectful and intolerant is the inevitable outcome of his relativisation of the principles of practical reason and justice. His context-dependent justificatory strategy implies that those outside that context can legitimately declare its principles inapplicable or unjustified to them. This puts Rawls in a paradoxical position: he is aware that liberal principles are (on his account) difficult if not impossible to justify outside the bounds of liberal societies, and yet he is also aware that an absolute relativism on the point of justice is not a satisfactory approach to international relations. He is therefore unable to detach worries about cultural imperialism from attempts to argue for justice beyond the borders of liberal regimes; his strategy forces him to choose between justice, in the full liberal sense, and respect, in the sense of non-imperialist cross-cultural relations. In other words, if the justification of justice is relative then we must choose between global justice and "respect" for other cultures.

With this in mind I would argue that his account of decent regimes can be described as an attempt to mitigate the starkness of these alternatives. It is an attempt to take both actual consent and its conditions seriously despite the fact that there now appears to be a wide gulf between what is consented to (a hierarchical nonliberal system) and what in principle can be consented to (principles grounded in freedom and equality). The account of decent regimes is therefore intended to describe hypothetical regimes whose systems of justice happen to fall between (context-dependent) liberal principles and (full) illegitimacy and as such are not properly the subjects of criticism from a liberal
perspective. Rawls is attempting to describe a scenario where liberal states do not need to choose between, on the one hand, a concern for justice beyond borders, and on the other, toleration and respect of other states. He is seeking to balance the need to avoid criticism, which on his account is necessarily intolerant, with a concern for the non-relativity of justice. In other words, the purpose is to describe a scenario in which the relativity of the justification of justice need not imply the absolute relativisation of justice itself.

I suggest that we may view the description of decent regimes as the method by which Rawls attempts to balance the competing and seemingly incompatible concerns of full liberal justice and toleration. The critical concept deployed by Rawls is that of "decency": decent regimes, while not endorsing full liberal justice, meet sufficient criteria with regard to political consultation and representation to claim at least limited legitimacy. As such the dissent procedures of the political system are crucial; they must be sufficiently robust to ground the claim that what is consented to by the members of such regimes, through loyalty or affirmation of the system of justice, may be considered legitimate, as opposed to deficient, spurious or coerced. It assures legitimacy by suggesting that what is consented to could have been dissent from. This is crucial because, as we noted above, it is consent legitimated through the dissent procedure that grounds the claim that there is agreement on, or affirmation of, a shared common good conception of justice (recall that Rawls described all members as affirming such an idea of justice). Hence these procedures ground any suggestion that decent regimes constitute a "people" with shared values who may collectively experience feelings of humiliation or wounded self-respect as a result of international criticism.

However, as I have argued, there is a circularity to Rawls's description of such procedures which implies that they fail to legitimate consent in any meaningful way; at the very least they fail to legitimate the claim that decent regimes as described constitute a people who share a conception of justice. The dissent procedure legitimises the agreement that grounds the shared nature of the conception of justice. However, the dissent procedure itself operates within the framework of this apparently shared conception of nonliberal justice. If the
conception of justice forms the framework for dissent then it is a presupposition of the dissent procedures and not subject to them. If it is not a candidate for dissent then any apparent consent to it cannot be considered legitimate; the legitimacy of consent is drastically undermined where the option to dissent is unavailable. Therefore an even partially legitimate consensus on the system of justice cannot be assumed.

Furthermore, we have also noted that the common good idea of justice is an aspect of the dominant comprehensive doctrine. It therefore seems more appropriate to assume the existence of a plurality of (both reasonable and unreasonable) perspectives on justice within decent regimes, rather than a common (legitimate) affirmation of a shared nonliberal conception of justice. It is quite possible then to imagine a legitimate affirmation of such a regime to be both unlikely and paradoxical, given the diversity of perspectives. This consideration further reinforces the possibility that there are no grounds to assume sufficient homogeneity within decent regimes to constitute a people worthy of the kind of “respect” that involves withholding criticism of potentially illegitimate political institutions.

What all this implies is that Rawls does not succeed in what he sets out to achieve: to describe a political regime as a “people” in the sense that the majority of its members legitimately affirm and share a nonliberal conception of justice. Instead, homogeneity is not a valid assumption, whereas the existence of a plurality of doctrines (some quite possibly reasonable) is a valid assumption.93 The

---

93 Rawls argued in Political Liberalism that only in a liberal democratic society can reasonable pluralism exist, because only under free institutions can a plurality of ethical doctrine arise. This raises a question with regard to decent societies: does it imply that reasonable doctrines cannot arise in the context of nonliberal institutions, or that they cannot flourish? The latter seems the more appropriate understanding, meaning that it is quite possible for reasonable comprehensive doctrines to be present in nonliberal societies. The political philosopher Andrew Kuper has critiqued Rawls’s account of decency on the basis that we must assume the existence of members of decent regimes who affirm principles of freedom and equality: “When a liberal regulatory framework recognises a decent hierarchical regime as sufficiently just, it participates in the denial of freedom and equality to such individuals. Dissenting individuals with liberal views would surely, it seems, dispute the idea that accommodation of reasonable pluralism requires that their individual moral claims be taken less seriously. But then one could not really know what they would think since their views could well be sealed off from view by the decent consultation hierarchy.” Andrew Kuper, "Rawlsian Global Justice: Beyond the Law of Peoples to a Cosmopolitan Law of Persons", Political Theory, Vol. 28, No. 5 (October 2000), pp. 640-674, p. 651.
hierarchical nature of the society and the dominance of a particular comprehensive doctrine drastically undermine the possibility of dissent and therefore of voluntary legitimate consent to political arrangements.

Rawls does not succeed in describing a regime whose institutions would not properly be the subject of criticism from the perspective of justice. This is not because these institutions do not replicate Rawls's liberal institutions, but because they do not meet the most basic conditions for assuming that those who live under the regime affirm their own social and political institutions in a morally significant sense. Furthermore, based on O'Neill's account of justice and agency (which we will examine in detail in Part 3), it cannot be assumed that those who live under such regimes can affirm their own tradition in any meaningful sense, because consent or affirmation depend on the protection of the underlying capacities to act that enable (and legitimate) acts such as consent or affirmation.

Therefore Rawls's account of decent regimes cannot mitigate the starkness of the alternatives presented as a result of his context-dependent justification of practical reason. No artful description of hypothetical regimes can override the difficulties in justifying regimes that deny principles of freedom and equality, as such principles protect the underlying capacities to act that allow for legitimate agreement.

We are left with the same problem: institutions that deny freedom and equality stand in need of criticism from the perspective of justice, yet, on Rawls's account of justice, there is no available (i.e. non-contextual) account of justice and practical reason with which to mount a critique that could not be dismissed as merely Eurocentric. In other words, decent regimes are only deficient from "our" liberal perspective, i.e. not from any objective perspective, hence his attempt to describe them as otherwise (as partially legitimate), and hence the paradoxical result: regimes that apparently receive legitimate endorsement despite denying the conditions that make such endorsement possible.
While Rawls’s aim of extending toleration is commendable, I have argued that the particular tension he identifies between “justice” and “toleration” derives from deficiencies in his own account that could be avoided by an alternative account of the justification of justice. The problems for his account, for example the status of decent regimes, arise from his insistence on taking consent (whether hypothetical or actual) as fundamental to justification; Rawls’s denial that institutions that fail to meet the basic conditions for legitimacy warrant criticism arises because he takes consent and not the conditions that legitimate it as the ultimate ground for justification. He is forced to take their consensus seriously despite its seeming illegitimacy. While he recognises that in the case of decent regimes there is a significant gulf between what can be accepted and what is accepted, his emphasis on probable consent in his earlier works now has the implication that what is actually consented to must be morally relevant.

What should be clear now is that it is Rawls’s own deficient and context-dependent strategy of justification that leads him to be suspicious of attempts to export or impose principles of freedom and equality on other nations or “peoples”. It is this strategy that leads him to begin his account of international justice with the premise that criticism of supposedly “decent” regimes amounts to a form of intolerance. Consensus forms the basis of justification for principles, therefore outside the consensus those principles have merely the status of partisan and relative principles. To attempt to criticise other societies for failing to embody principles which are only justified within one’s own society and are hence contextual is little more than cultural imperialism.

Rawls’s approach to international justice underlines the inadequacy of his strategy of justification and highlights the need for a universally justified account of practical reason. Only an account that could be shared by all (could, not is, or would be) can ground the necessary criticism of unjust or illegitimate regimes. Without such an account we are faced with little choice but to accept the relativity of reason and renounce the possibility of identifying an objective perspective from which to critique institutions that fail to meet standards of justice.
2.2.5. A shared perspective on international justice? Decent regimes as agents in the original position

As was noted above, the description of decent regimes is the method by which Rawls attempts to overcome the context-dependent nature of the justification of liberal justice and political institutions in his account. It is an attempt to avoid the imposition of particular political values upon distinct (and acceptable) others whilst also not falling below a minimum standard of justice internationally. A primary concern for Rawls here is stability; the forceful pursuit of “liberal” justice would lead to an unstable international regime. The concerns of stability and justice must be balanced, and this means conceding that full liberal justice is not attainable in an international context made up of diverse ethical perspectives and political regimes, i.e. agreement on full liberal justice is not attainable. The question then for Rawls was how to promote both stability and justice in a world where others disagree on the content of political justice?

While the first aspect of his solution was conceiving of a society that denied principles of freedom and equality yet (paradoxically) could also be viewed as a homogenous people worthy of due respect as such, the second stage involves the construction of a hypothetical scenario in which such societies agree to affirm a law of peoples that allows them to coexist with “us”, liberal societies. If such societies affirm the law of peoples, which has already been affirmed by liberal societies, then this consensus can form the basis of public justification, i.e. the account of public reason in the international context. In this way “urgent” human rights, a minimal subset of liberal rights, are grounded and justifiable, as are the rules governing the conduct of states. Therefore both stability (agreement and respect) and justice (human rights and at least partially legitimate regimes) have been ensured. The nature of decent regimes is critical to this account. Only states that refrain from criticising each other can share in a stable Society of Peoples; criticism on Rawls’s account is intolerance and leads to disagreement, conflict and

---

94 This is the third original position. The second original position involves only liberal societies agreeing to share the law of peoples. I do not analyse the second original position separately, rather I will address it in relation to our analysis of the third.
Practical Reasoning and Transnational Justice

instability. Cooperation and critical engagement are mutually exclusive on Rawls's account and therefore decent regimes must be beyond criticism.

Once acceptable regimes (liberal and decent) have been hypothesised, the hypothetical procedure by which they select principles for international law can be set out. This procedure replicates the original position in the domestic case; it models the appropriate conditions, i.e. it situates all the relevant agents in a relation of equality with each other and deprives them of information that may lead them to select principles that benefit them to the disadvantage of other, possibly weaker, agents. While much remains the same, the crucial difference is that the agent in question is now the state acting on behalf of its members, i.e. it is peoples and not persons who choose the basic principles of international law.

Rawls's defence of the focus on peoples and not persons in *The Law of Peoples*, or on the role of boundaries, takes up the objection that fixed state boundaries are arbitrary and therefore cannot be justified within a *Society of Peoples*, in particular as the relevant agents in the original position. He argues that "to fix on their arbitrariness is to fix on the wrong thing", because despite their arbitrariness they are necessary features of a global context, i.e. "[i]n the absence of a world-state, there must be boundaries of some kind" (LP, 39). Furthermore, we need not dismantle boundaries to achieve a "reasonably just (or at least decent) Society of Peoples", rather the limits on "inequalities of wealth and power" can be managed by considering what "reasonably just liberal peoples and decent peoples owe to societies burdened by unfavourable conditions" (LP, 39). In other words, questions of global economic justice (which for Rawls centre around the Duty of Assistance both for liberal and by decent regimes) can be settled after questions of domestic justice have been addressed.

Of course, there is a further consideration that Rawls does not address here that is crucial to his strategy for justifying why the perspective of peoples over individuals is taken. The practical necessity of boundaries does not actually explain why we must take the *state* or *people* as the relevant agent for the selection of the principles of international law (and therefore international justice) that will
dictate the lives of the persons living within and beyond those states. An obvious concern is that some persons or groups may be excluded and their interests not represented or considered in the selection of principles. This would not be an acceptable outcome for Rawls, and he does not think his account runs this risk. To understand why, we must review the nature of peoples, both liberal and decent, in Rawls's account. I will first explain, from Rawls's perspective, how the nature of peoples implies that they can legitimately be taken as the appropriate agents in the original position, before revisiting some of the difficulties with this account to highlight the questionability of this claim.

Taking Rawls's description of a just liberal society first, while there exists a diversity of reasonable ethical perspectives in such societies, there is also an overlapping consensus which forms the basis of public reason; there is a shared agreement, *legitimated by the principle of reciprocity*, on the basic principles of political justice. This implies a *neutralisation* of the diversity of ethical perspectives resulting in a *homogenisation* of the perspective on the political conception of justice in the public sphere. This homogenisation implies that, at the level of public reason, *any* citizen can be taken to represent the perspective of *each*. This is due to their sharing in a conception of justice, which is legitimated through the principles of reciprocity, i.e. by the fact that they are situated equally in relation to each other. A people, as a collectivity united around a shared conception of justice (along with a sense of shared identity and common sympathies), is constituted by persons whose perspectives in the public sphere are interchangeable, this is what is meant by a homogenous people with a shared conception of justice. Furthermore, if the state is the legitimate representative of the people, then the state can be taken to legitimately represent the perspective of the individual agents within the regime.

In the case of decent regimes, these too, according to Rawls, can be taken as peoples who share a perspective on justice in their public sphere. The dissent procedures are sufficiently robust to ensure that less equal members are not coerced or forced to submit to the regime, and on this basis their loyalty toward the regime and their common affirmation of a nonliberal conception of justice can
be taken as (at least a partially) legitimate form of consent toward the regime. As in the case of liberal societies, the shared nature of the account of nonliberal justice ensures that there is a sufficient level of homogeneity in the public sphere for such a regime to constitute a people. In this way, as with liberal societies, the perspective of any member (in the public sphere) can be taken as the perspective of each. And again as with liberal societies, a people is constituted by persons whose perspectives are interchangeable in the public sphere and the state can be taken to legitimately represent the perspective of the individual agents within the regime.

In both cases it is the sharing of a perspective on justice in the public sphere that justifies taking a people over persons as the relevant agent; homogenisation of diverse perspectives implies that the perspectives of individual human agents are all (indirectly) accounted for at the level of the original position between peoples. While in liberal regimes the assumption of homogeneity at this level is justified with reference to the fulfilment of a criterion of reciprocity that ensures principles of fairness form the basis of the consensus, thereby ensuring agreement among reasonable persons, in decent regimes it is the sincere and realistic option of dissent that ensures the legitimacy of any apparent consent to the regime of nonliberal justice.

However, as we have seen, there are significant difficulties with this line of argumentation with regard to decent regimes. Recall that dissent “is permissible provided it stays within the framework of the common good idea of justice” (LP, 72). The conception of nonliberal justice is assumed within the framework of the dissent procedure and therefore cannot be an object of dissent. If it cannot be an object of dissent then it is questionable that any apparent consent to such a conception is even partially legitimate. If there is no impartial framework to manage objections to the conception of justice itself then any apparent loyalty or affirmation of the regime cannot be presumed to justify the assumption that the diversity of ethical perspectives has been neutralised in the public sphere. Freedom and equality are essential preconditions for the legitimacy of consent and without these conditions it cannot be assumed that consent has been given. In the
case of decent regimes homogeneity is not the appropriate assumption, rather the plurality of doctrines must be taken as just that, a plurality of ethical perspectives on justice. Without a legitimate consensus on the conception of justice the decent regime cannot be taken to constitute a people. In this case it cannot be assumed that the perspective of any member can be taken as the perspective of each, and therefore the state, insofar as it purports to represent a people, cannot be taken as the relevant agent for reasoning about the principles of international justice that will be enforced upon its members.

Furthermore, it can be assumed that not only do a diversity of perspectives exist but they exist in a hierarchy whereby it is the dominant doctrine whose conception of nonliberal justice guides the regime. It is this group whose perspective constitutes the relevant agent for the selection of principles at the international level, implying that the interests and needs of the subordinate group/s risk exclusion from consideration. Rawls may respond that the ruling group take the perspectives of their members seriously, however I have argued that without an independent or impartial framework for dissent there is no guarantee that the procedures of consultation are effective. If there is no such guarantee, then there remains a tangible risk that some perspectives will be excluded.95

These are of course theoretical considerations; an actual original position does not take place. Their point is to highlight that if states are taken as the relevant agents in the international original position, then the principles of international justice are not justified to all the members of the societies that make up the Society of Peoples, in particular to the members of subordinate groups within decent regimes. The consensus that is hypothesised to emerge could not be considered

95 The political philosopher Simon Caney has also critiqued the consultation procedures of decent regimes on the basis that they cannot guarantee that voices will be heard. He argues that the “right to express dissent” may not be effective in changing the minds of leaders, and also that a right to merely express dissent is no guarantee of a right to actively dissent. That judges explain their reasoning does not imply that repressive policies will not still be enacted; the right to express dissent is purely formal and does not preclude unjust policies. Simon Caney, “Cosmopolitanism and the Law of Peoples”, in John Rawls: Critical Assessments of Leading Political Philosophers, ed., Chandran Kukathas (London: Routledge, 2003), pp. 263-92, p. 272.
legitimate and could not function as a universally acceptable account of public reason in the international context of the “Society of Peoples”.

Rawls of course argued that the lack of equality within decent societies was not a sufficient reason not to treat them as equals in the original position and the Society of Peoples. He claimed that “equality holds between reasonable or decent, and rational individuals or collectives of various kinds when the relation of equality between them is appropriate to the case at hand”, e.g., although church organisations are structured hierarchically, the separate “churches may be treated equally and are to be consulted as equals on policy matter” (LP, 69). Regimes should be treated in the same way as churches in this regard; their internal inequalities do not imply that they are not each due equal consideration in matters of relations between regimes.

This defence of course only makes the claim that domestic inequality is an insufficient reason to deny regime equality. If we look further back though at why liberal societies ought to treat decent regimes as equals in a Society of Peoples, we see that the claim that decent regimes ought to be treated as equals relates to the idea of the self-respect of a people, which may be “wounded” by the denial of respect from other peoples. Therefore, once again, the claim and the defence rest on the assumption that we may conceive of decent regimes as a people. This assumption does not however hold; there is sufficient diversity to dismiss such an assumption and to assume instead a variety of potentially reasonable perspectives. The decent regime is not sufficiently homogenous to ground the claim that its members are a collective unit capable of experiencing respect or disrespect in any legitimate or justifiable sense.

Furthermore, the comparison with church organisations only reinforces the point that decent regimes deny the basic conditions for even partial legitimacy. The crucial distinction between church organisations (assuming we are referring to

96 “Recall that peoples (as opposed to states) have a definite moral nature... This nature includes a proper pride and sense of honor; people may take a proper pride in their histories and achievements, as what I call a 'proper patriotism' allows... The due respect they ask for is a due respect consistent with the equality of all peoples”. (LP, 62)
those within liberal societies) and decent regimes is that the church does not deny the basic freedom and equality of its members. In fact it does not have this power, for within liberal states it is the political institutions that guarantee freedom and equality and circumscribe religious and other institutions within these limits. Therefore, any consent to be a member of a church that is structured hierarchically can in principle be considered legitimate because the conditions that ensure its legitimacy are met in the broader context within which the church operates. The basic freedom of association of persons is not undermined or denied by their belonging to a particular religious organisation. At a basic moral and political level the members of the church are still considered free and equal moral persons and therefore the separate churches may be treated as equals. The same cannot be said of members of decent regimes. Their actual membership in such societies is (most likely) a matter of birth and not of consent or choice, there is (on Rawls’s account) no wider context that assures them freedom of choice and equal basic rights in this regard, and they therefore cannot be thought of as having agreed to be a member of a hierarchical system in the same sense as church members.97

Equality between states, with the implication that their regimes are respected and hence seen as beyond criticism, only holds where it is clear that the shared conception of justice is in fact so, i.e. where the conditions for legitimate consent have been met. Decent regimes as hypothesised by Rawls fail in this regard and as

97 These considerations are compounded by Rawls’s thinking on immigration and emigration. Of rights of migration, Rawls argues that a people, as responsible for a territory, cannot make up for neglect of that territory by migrating elsewhere (LP, 8-9). He later adds that “[t]his remark implies that a people has at least a qualified right to limit immigration. I leave aside here what these qualifications are. There are also important assumptions I make here which are not considered until Part III... where I examine the duties of well-ordered societies to those societies burdened by unfavorable conditions. Another reason for limiting immigration is to protect a people’s political culture and its constitutional principles” (LP, 39). However, later, he asserts that “in view of the possible inequality of religious freedom, if for no other reason, it is essential that a hierarchical society allow and provide assistance for the right of emigration” (LP, 74). Leaving aside some obvious concerns, such as the merits of a strategy of cultural isolation in a globally integrated world, or whether a “right” of emigration is any right at all, I note that the two comments seem mutually incompatible. On a very straightforward assessment, it appears that even a qualified right to limit immigration must cancel out any rights of emigration. The “option” to leave is meaningless when there is nowhere to go, or, as the political philosopher Kok-Chor Tan asks, “is giving one the right to leave one’s country a real choice?” Kok-Chor Tan, “Liberal Toleration in Rawls’s Law of Peoples”, in John Rawls: Critical Assessments of Leading Political Philosophers, ed., Chandran Kukathas (London: Routledge, 2003), pp. 193-211, p. 205.
such their agreement on the law of peoples cannot be viewed as an agreement between equal agents and cannot ground any claims as to its legitimacy. Agreement in the third original position fails to meet basic conditions for legitimacy; it cannot be presumed to ground an objective perspective. This is one sense in which the state cannot be taken as the appropriate agent.

However, even if it could be shown that decent regimes could meet such criteria, the implication of our earlier discussion of Rawls's deficient strategy of justification is that Rawls's entire account of legitimation and objectivity through agreement cannot ground an objective perspective. What this means is that while in the case of decent regimes the state as non-representative of the people cannot be taken as the appropriate agent, there are other more fundamental considerations that question whether the state (even in a society of peoples made up only of liberal democracies) is the appropriate starting point for an account of international justice. In the next two sections I examine the cosmopolitan critiques of Rawls's account of justice that ask whether justice ends at state boundaries.

2.3. The justification of state boundaries: exclusion and obligation in the context of global and transnational interaction

The previous section concluded that decent regimes could not be presumed to legitimately share a conception of justice in a manner sufficient to constitute a people that could function as a collective agent at the international level. This does not, however, imply that an international order of purely liberal regimes could succeed in this regard, rather there are further considerations, implied by our critique of the consensus-based justificatory strategy deployed by Rawls, that prevent even a system of liberal states from being the appropriate starting point for a theory of international justice. While in Rawls's account of political liberalism the gulf between possible and probable consent was obscured by the apparent correspondence between what could be agreed to and what, it was hypothesised, would in fact be agreed to, it is now clear that probable consent is the basis of Rawls's account of justification. Probable consent (in the form of actual or
hypothetical consent) is not capable of grounding an objective perspective, i.e. it cannot justify the conditions for its own legitimacy. I now argue that if the consensus cannot be taken as legitimate or objective, correspondingly the territorial boundary that coincides with the consensus can also not be automatically taken as legitimate. Rather the exclusion that is implied by such a boundary must be justified to those within and outside of the consensus, i.e. it must also be justified to those who are excluded. Such a justification can only be offered by a perspective that is not grounded in the consensus, because it is this consensus that is in need of justification and this exclusion that is in need of critique (2.3.1.).

This claim is contingent on the assumption that those outside the consensus are morally relevant others. With this in mind I turn to consider Charles Beitz’s and Thomas Pogge’s immanent critiques of Rawls’s state-orientated account of global justice. Where Rawls views as morally relevant those others with whom persons share a basic political structure (and the common sympathies and shared values that generates), Pogge and Beitz argue that persons interact with others beyond our borders in ways that have deep and lasting effects on agency, and as such make us morally accountable to those others. Beitz argues that the appropriate view of the world itself is as a powerful and non-voluntary scheme of interaction. Pogge also argues for such a perspective, and specifies that the system creates unreasonable inequalities that have inescapable and sometimes harmful effects on individuals’ agency. He also argues, drawing on the work of Amartya Sen, that the form of impartiality deployed by Rawls in his account is subjective from the perspective of outsiders; it is only “closed impartiality”. “Open impartiality” would involve the inclusion of perspectives from outside the domestic society, or the consensus, as we have discussed in our critique of Rawls’s reliance on the internal liberal perspective. Pogge also argues that our participation in the global system makes us complicit in its effects. He claims that we are all obligated to either reform the system or compensate for its injustices, based on a negative duty to avoid harming others. Given that our obligations toward distant others are therefore a matter of duties of justice, he rejects the duty of assistance as inadequate. Assistance is not the appropriate category for assessing global poverty or global justice. The power and vulnerability generated by the global web of
interaction are the appropriate considerations of a theory of international justice, rather than reasoning from the point of view of a particular liberal society (2.3.2.).

Finally, I address a potential objection to this critique: that the duty of assistance applies alongside a fair global economic system. In response I note that the duty of assistance, as a feature of nonideal theory, must be applicable to the actual circumstances of injustice. I argue, again drawing on the work of O'Neill, that Rawls in fact introduces idealised premises into his account of burdened societies, and as such undermines the relevance of the proposed duty of assistance to actual underdeveloped societies. The introduction of ideal premises, specifically the assumption that burdened states exist against the background of a fair global economic system, obscures the actual duties that arise based on the nonideal circumstances in which we find ourselves. I conclude the duty of assistance is an idealised concept only applicable to idealised burdened societies. Furthermore, it is incompatible with the actual duties of justice; by making assistance the central duty of liberal persons, it denies that more concrete duties of justice exist toward distant others (2.3.3.) These duties will have to be further specified so as to prevent the recasting of duty as a matter of optional, discretionary benevolence or beneficence.

2.3.1. Objectivity and boundaries: the status of a consensus in relation to justifying exclusion

In the previous section we saw that an examination of "decent" regimes demonstrates not only that consent in such regimes does not meet conditions of legitimacy, but also (based on O'Neill's critique) that consent itself is not fundamental to justification. The case of decent regimes highlighted that the correspondence between possible and probable consent in political liberalism was by no means guaranteed. It also made clear that the emphasis for Rawls was always on what would be agreed; while Rawls noted that certain conditions must be met before consensual acts could give rise to obligations, i.e. be considered legitimate, in the end his justification of these conditions relied on consent, i.e. on
the actual endorsement of the audience of justification. Consequently he was forced to treat the apparent consent of members of decent regimes as an indication of legitimacy.

Rawls did not succeed in justifying the conditions that were necessary to lend objectivity to any endorsement of his account of justice. Following the work of O'Neill, I concluded that Rawls's account could not do without an account of justification that argued for its premises, and avoided relativism by grounding reasons for the objectivity of the overlapping consensus that formed the basis of public reason. Now in the case of the international original position(s) this objection is extended. The account of public reason, whether domestic or international, cannot ground an objective account of justice or a law of peoples because it relies in the end entirely upon the perspectives of those within the consensus, whether a consensus of liberal citizens within a shared polity or a consensus of liberal and decent peoples. The case of the international original position(s) faces the same problems of justification as in Political Liberalism. Reliance on consensus means relying on a set of shared premises. In the case of the bounded society (as discussed in A Theory of Justice and Political Liberalism) there is first a consensus on the appropriate conditions that ought to influence the selection of principles of justice, which is followed by a confirmation through considered convictions, and in political liberalism by an overlapping consensus, on the selected principles of justice, which forms the basis of public reason. These steps are repeated in the analogous case of the international original position(s), first with a consensus on the situating of states in a position of equality to each other, and consequently with an overlapping consensus on the principles of international justice which form the account of international public reason, i.e. the law of peoples.

Rawls's strategy of justification, by invoking a consensus as a premise, invokes a boundary as a fundamental aspect of justification. In the case of the domestic consensus this can be taken to imply the actual territorial boundary, or more precisely it implies the delineation of a consensus on certain principles that grounds a system of cooperation and coincides with that territory. And just as a
consensus cannot justify the conditions of its own legitimation, a boundary (which
delineates the consensus) cannot be presupposed in its own justification, i.e. in the
justification of the exclusive political institutions of a specific bounded society. If
we cannot take consent or consensus as fundamental, then we cannot take the
bounded society (which according to Rawls is grounded in a consensus on certain
principles) as fundamental in the sense of being prior to questions of justice. We
cannot look primarily to the fact of agreement around a set of political institutions
as justification for their existence.

This of course does not imply that boundaries are irrelevant, rather that, as with
consent, there are prior considerations that ground a shared agreement as
legitimate from an objective or impartial perspective, thereby preventing it from
being considered merely partial and, at worst, arbitrary. As we have seen, an
actual consensus does not merely fail to justify to those outside the consensus:
alone it offers no justification at all. A consensus cannot serve as an objective
perspective regardless of one's position in relation to the consensus. However,
from an internal perspective the partial nature of the justification of political
justice is more likely to be obscured, due to the fact that it coincides with or reflects
what is actually wanted or chosen. It is more likely to be identified as partisan and
unjustified by those outside the consensus when the particular conception of
justice, and the system of rights and duties it enshrines, comes into conflict with
the needs or aspirations of those outside the consensus, i.e. when questions of
exclusion, human rights or distributive justice arise.  

This is why we may speak of justification as denied to those outside the consensus,
when more accurately justification is also absent within the consensus. In fact the
very act of distinguishing between the internal and external perspective reminds
us that it is this very division that stands in need of justification; if there are

98 It must also be noted that this is a simplified presentation: questions of exclusion, conflict over
identity, ethical pluralism and distributive justice arise both within and across so-called bounded
societies, the delineation of those within and outside a consensus does not generally reflect
national boundaries, a point which O'Neill makes when presenting her account of public reason.
For now I examine only the division between those inside and outside national boundaries, as it is
the justification of state boundaries in relation to questions of international justice that is our
concern here. We will discuss internal rejection of the consensus when we turn to alternative
approaches to public reason in Part 3.
considerations prior to the agreement that must be taken into account, such as the conditions that legitimate it (which involves the justification of action to relevant agents), then these must necessarily apply prior to the partitioning of perspectives between those inside and outside the consensus. It would be incoherent to suggest that the considerations that legitimate consent need only apply within the bounded society, i.e. within the consensus, as the internal-external distinction is a feature of the consensus that awaits justification. Such a suggestion would amount to presupposing agreement or consent as its own justification.

What this implies is that exclusion, i.e. the internal-external distinction, must be justified from an objective (hence not biased or partial) perspective, in the same sense that the distinction between the reasonable and unreasonable must be grounded in an objective perspective. As in that case, this perspective must necessarily be independent of a particular shared perspective. The boundary itself cannot justify exclusion; the consensus, i.e. the agreement between certain persons on what principles, rights and duties they would endorse, cannot suffice as justification for the exclusion of others.

When we discuss the justification of exclusion we are, of course, also discussing the justification of political institutions, because it is not exclusion itself that Rawls seeks to justify, but the political conception of justice for a bounded society. The bounded society, by its nature, implies exclusion, and if those who are excluded are owed justification, then such exclusion must be justified from a perspective that they also can accept, which I have argued cannot be the perspective of those who seek to exclude. The price of objectivity is the recognition that the domestic system of rights and duties does not exist in a radically isolated sphere of interaction, but rather in the context of a shared sphere of interaction with non-domestic others to whom justification must be extended.
2.3.2. Obligations to distant others: immanent critiques of Rawls’s state-centred account of international justice

I have argued that an objective account of the justice of political institutions must be capable of justifying itself to those who are excluded. There are two respects in which we may speak about exclusion from the consensus; in terms of those rejected as unreasonable and those excluded as irrelevant, i.e. as non-members of the relevant bounded society. While I noted that the cases were similar (both types appear to be excluded on grounds not justified to them) there is also a critical difference. In the case of exclusion on the basis of unreasonableness, the objection that this cannot be justified via agreement is relatively complete (and will be further considered in Part 3 on public reason). However, in the case of exclusion on the basis of irrelevance, the objection that this has not been justified to those others is contingent upon a further consideration that relates to “the scope of moral concern”. \(^{99}\) The crucial point is that exclusion of those outside the bounded society can only be claimed as unjustified (meaning that those others are owed justification) if it can also be shown that they are within the scope of moral concern. If they are not, then they are morally irrelevant and their exclusion is no problem for justification. It is this point that I will now treat, as it is Rawls’s thinking on this subject that influences his account of “our” obligations to distant others, especially with regard to economic or distributive justice.

Since exclusion is only problematic where the other is a morally relevant subject or agent, a theory of justice must include a method for identifying which persons fall under this category. Rawls views non-members of the domestic society as moral persons, however he does not see them as relevant to the justification of the basic structure of a bounded society, which is the primary focus of his account of justice. Given that he assumes that citizens of liberal societies have only very limited interaction with others, most of which is mediated through national and international institutions, he does not see such interaction as relevant to the justification of basic political institutions. In other words, he does not think these

interactions have any bearing on the establishment or operation of a legal framework of civic rights and duties. Agreement, as we have seen, is fundamental to his account of justification, and in this way the relevant form of interaction for defining the moral horizon is living under shared political institutions. Such a framework institutionalises the principle of reciprocity that he believes makes persons accountable to one another: “those who can give justice are owed justice” (TJ, 510).

It is not that outsiders, non-domestic agents, do not possess the “two moral powers” that make them capable of giving and receiving justice and therefore qualifying them as moral persons, but rather that within a bounded society persons’ actions towards one another are mediated through the basic political structure; it is the mediation of others’ action and transactions through shared political institutions that makes those others morally relevant, as it is these institutions that specify our rights and duties, hence the distribution of primary goods, and hence our capabilities and opportunities. This basic system of rights and duties that regulates our opportunities is of fundamental and primary concern, therefore from the social contract perspective it is interactions that relate to this system, i.e. to the process of consensus-forming on this structure, are most relevant to justice. Therefore only those involved in the formation and ongoing realisation of the social contract count as relevant others. Domestic citizens only interact in relevant ways with external others when they establish the laws of international relations, i.e. the legal framework to govern the space in which states interact, and this of course is drastically restricted by comparison, because, according to Rawls, most matters of justice between persons are handled domestically.

For Rawls it is the system of rights and duties that protects our capabilities and ensures our opportunities; action or interaction that is not relevant to this system is not relevant to justice “as a virtue of social institutions” (TJ, 3) (although they are to the sense of justice). Crucially, this implies that non-domestic interaction has little or no bearing on the establishment or justification of rights or duties, i.e. citizens of the bounded society do not, and cannot, have obligations of justice.
Practical Reasoning and Transnational Justice

(beyond respect for the law of peoples) toward those outside their particular bounded society.

Rawls's focus on the basic structure then is guided by the most important concern of justice: the protection and promotion of capabilities and opportunities, i.e. of agency. Where he differs from his critics then is not simply that he makes an arbitrary distinction between domestic and international institutions. The crucial distinction is that where Rawls sees the important influences on agency as restricted to domestic institutions, his critics hold that agency is also influenced by the distribution of benefits and burdens transnationally. For them, much of the action relevant to justice is not mediated through domestic political institutions, and hence there is scope for the idea that rights and duties may arise toward others with whom we do not share domestic political institutions.

Along with O'Neill, Charles Beitz and Thomas Pogge are two notable critics who have argued that the international or global system affects individual lives in ways deeply relevant to questions of basic justice. I will review some of their objections to the distinction between domestic and international justice and the implications for the possibility of duties of justice across borders. I will also comment on similarities with O'Neill's critique, before turning to her assessment of the fundamental inadequacy of a conception of "aid" or "assistance" when considering problems such as global material or distributive justice.

Charles Beitz argues, on grounds of political reality, that assumptions of self-sufficiency such as those deployed by Rawls in *The Law of Peoples* are implausible, that empirical evidence shows that the world itself has become a scheme of social cooperation. He cites many examples of increasing global interdependence; particularly he notes that the ascension of the international division of labour, whereby labour, price and wage are controlled by multinational entities, means "value created in one society is used to benefit members of other societies".¹⁰⁰ He also notes how the global economic system has developed its own financial and

monetary institutions.\textsuperscript{101} This international order, he argues, imposes inescapable burdens on the least advantaged nations. He outlines three of the most devastating burdens; 1) adverse balance of payments scenarios for developing societies often require them to sell to foreign interests resources that could potentially be used for development at home, 2) there is a definite loss of political autonomy evident where poorer nations must participate in the global economy "on the only terms available", and 3) the global monetary system provides few safeguards against the global implications of price inflation in wealthier societies.\textsuperscript{102} The migratory nature of profits and burdens internationally leads Beitz to dismiss as implausible the suggestion that a nation's fate is dependent upon predominantly internal factors, and to suggest that "we should not view national boundaries as having fundamental moral significance – since boundaries are not co-extensive with the scope of social cooperation, they do not mark the limits of social obligations".\textsuperscript{103} Instead, the levels of global social and economic interaction which have been sought and partially achieved, since at the very least the beginning of the post-war period, entail a moral requirement to extend the horizons of our moral obligation.

Furthermore, he observes that the question of whether a society's burdens are the result of internal or external factors may itself be an opaque formulation; he writes:

\[\text{[I]t is not even clear that the question is intelligible as it arises for contemporary developing societies enmeshed in the global division of labor: a society's integration into the world economy... can have deep and lasting consequences for the domestic economic and political structure. Under these circumstances it may not even be possible to distinguish between domestic and international influences on a society's economic condition.}\textsuperscript{104}\]

\textsuperscript{101} Beitz, "International Relations", p. 296.
\textsuperscript{102} Beitz, "International Relations", p. 296. All three, and many others, have been comprehensively documented by reports from the development sector, and by notable academics, such as the Nobel Prize economist Joseph Stiglitz, in his book, Making Globalization Work, and its predecessor, Globalization and its Discontents. (See Joseph Stiglitz, Making Globalization Work (London: Allen Lane, 2006), and Joseph Stiglitz, Globalization and its Discontents (London: Allen Lane, 2002). This inescapable and potentially burdensome interdependence is also demonstrated by the effect of the crash of the financial system from 2007 onwards on areas beyond North America and Europe.
\textsuperscript{103} Beitz, "International Relations", p. 298.
We begin to see self-sufficiency or autarky as a problematic premise to employ even in hypothetical and abstract scenarios relating to social development; the global economic matrix within which all societies must function to some extent, may not be something that can be bracketed when reflecting on the establishment of principles necessary for the development of "well-ordered" societies. The arguments from self-sufficiency then appear not to hold and it seems there are good reasons to avoid employing such an assumption.

Beitz also suggests that "mutual cooperation" may itself be an inadequate criterion by which to predicate issues of domestic or global justice. He focuses on the particular character of global economic relations, namely that global economic interdependence involves a pattern of relationships that are largely non-voluntary. He suggests that we should conceive of the overarching global structures as "powerful" and "nonvoluntary", and judge social and economic interactions according to their effect on the well-being of affected persons.\(^{105}\)

In summary, what Beitz is aiming to draw attention to with his critique of The Law of Peoples is what he terms "the instability of a theory resting on the sharp distinctions between the domestic and international realm".\(^{106}\) He also observes that Rawls's approaches to human rights and international distributive justice may be directly connected to his tendency to view states as moral agents.\(^{107}\) He points to potential ethical and practical, political concerns arising from the arbitrary location of natural resources, and to deficiencies in theories predicated on assumptions of self-sufficiency that do not adequately account for the extent to which individual societies are dependent on the larger schemes, whether multilateral, federal or global, within which they find themselves.

It is Thomas Pogge's fundamental assertion that the priorities of Rawls's own theory of justice demand that we always take the fundamental interests of persons over against other interests. He argues for this internal critique by drawing attention to the potential inadequacy of state-orientated theories in coping with

---

105 Beitz, "International Relations", p. 304.
the complex demands of equality, to the danger of premising a system of global justice on social structures that are only minor components of a larger comprehensive framework of action and interaction, and to the incompleteness of decision-making processes that exclude some whose voices should be heard.

Just as in the domestic society, background justice is an essential component of any global scheme of interaction. Social systems can create unreasonable inequalities, and Pogge observes that this is a “moral defect in a social system because many of the voluntary interactions taking place within will be morally flawed as a result”. In order for international relations to be genuinely free and fair, and free from coercion, there must exist “global background justice”. Pogge writes that “the grounds on which Rawls holds that fair equality of opportunity and the difference principle constitute requirements of background justice militate against confining these requirements within borders”.

Pogge also rejects the attempt to salvage the state-structured original position, which suggests such an original position would work were it to allow the representatives of states to take into account of differentials in power and resources. For Pogge this is still inadequate, because as long as we treat inequality between members of the same society and inequalities between societies separately, we allow the potential for severe equality gaps to develop between persons from different nations. Speaking of this dual criterion, he writes: “one observes that it would provide an incentive to justify otherwise excessive interpersonal inequalities... through the interposition of national borders.”

Rawls’s account does not seem to allow for the possibility that vast differentials in

109 Pogge, Realizing Rawls, p. 248.
111 Pogge, Realizing Rawls, p. 250.
112 Pogge, Realizing Rawls, p. 253.
113 Pogge, Realizing Rawls, p. 253.
wealth, regardless of whether a baseline has been established, could encourage less than free and fair processes in both domestic and international affairs.113

Like Beitz, Pogge also addresses some of the practical features of de facto international justice that are overlooked by Rawls's thesis, such as the non-voluntary nature of certain global arrangements, expressed by Pogge as an inescapable institutional order to which all are vulnerable, yet from which some lack the effective capacity to protect themselves. Pogge holds that Rawls would agree with Kant that just institutions are necessary "among human beings... who cannot avoid mutually influencing each other"114, and so if it can be shown that the world at large is a system of interaction, just institutions are required, too, at a global level. He quotes Rawls on the significance of the background structure in the domestic case, suggesting the same reasoning can be applied globally; agents "cannot comprehend the ramifications of their particular actions viewed collectively, nor can they be expected to foresee future circumstances that shape and transform present tendencies".115 All schemes of interaction possess a basic structure, and if it can be shown, as Rawls has attempted to do, that such a structure should be fair in the interests of justice, then it is a requirement of justice that the fairness of the basic structure be a criterion for the justice of any scheme of interaction.

Pogge writes that, by failing to assess the effects of domestic institutional structures on those outside of national borders,

\[w\]e disregard the negative externalities a national contract may impose upon those who are not parties to it... the criterion of domestic justice must then be adopted from a point of view that combines both perspectives from a suitably constrained standpoint of persons both insiders and outsiders... Such a standpoint is afforded by the global original position.116

113 In the next section (2.3.3) we will see how this question, of the harm caused by differentials in power, is also taken up by O'Neill in her discussion of coercion.
114 Pogge, Realizing Rawls, p. 241.
115 Pogge, Realizing Rawls, p. 255.
116 Pogge, Realizing Rawls, pp. 256-257.
This observation reflects our earlier consideration of the need for a properly objective perspective to ground justification, i.e. a perspective not grounded in the consensus. This is also reflected in a similar criticism raised by the economist and Nobel Laureate Amartya Sen, a criticism taken up by Pogge in an article, written jointly with Mark Labude, entitled “The Idea of Justice from a Rawlsian Perspective”.117 Sen suggests that there are two forms of impartiality, “open impartiality” involving the inclusion of all relevant perspectives, and “closed impartiality” involving only the perspectives of the particular “focal group” partaking in decision-making processes.118 He invokes Adam Smith’s “impartial spectator” to illustrate how the Rawlsian original position exercises only “closed impartiality”. The “impartial spectator” would require that neutral persons from outside the domestic society be involved in assessing the just formation of the basic structure, and in this sense he suggests it is a more demanding and universal conception of fair judgment than that of the original position.119 Sen notes that Rawls suggested in A Theory of Justice that the “impartial spectator” could be interpreted as an “ideal observer” compatible with justice as fairness. However, Sen observes that this is only one interpretation, and not Smith’s.120

With regard to global justice, Sen makes the observation that international relations account for only part of the ways in which people interact throughout the world and beyond borders. One may be led then to conclude that an adequate theory of global justice would require more mechanisms than only international agreements on principles of just relations, as these may not be capable of taking into account all relevant relations between persons.


119 Sen, “Impartiality”, p. 57. This critique, with its distinction between open and closed impartiality, parallels O’Neill’s Kantian distinction between public and private reason, in that Rawls’s account of public reason can be identified with “closed impartiality” on Sen’s account, and with private reason on O’Neill’s Kantian account, which we will discuss in Part 3.

120 Sen, “Impartiality”, p. 59. Sen also addresses the question of whether this form of “open impartiality” across borders could sustain the “public frameworks of thought” that are so essential to Rawls’s theory. In response he first quotes David Hume’s assertion that “the boundaries of justice grow larger in proportion to men’s views”, before noting that “the possibility of communication and cognizance across the borders should be no more absurd today than it was in Smith’s eighteenth century world.” Sen, “Impartiality”, p. 69.
Pogge observes that Rawls insists, in the domestic case, that all members of a society must be able to accept the principles of the basic structure, as this structure will govern their legitimate expectations for how they could or would choose to live their lives, and will be an influencing and inescapable force from the very beginning of life.\textsuperscript{121} Pogge argues analogously that any domestic society must take into account all those who will be affected when establishing its basic principles. Here he argues that Sen has correctly noted that the Rawlsian original position may be required to include the interests of individuals not belonging to a given polity, but who may be “profoundly affected” by its basic structure.\textsuperscript{122} If a given society fails to take into account the perspectives of outsiders, they are then invoking only a form of “closed” impartiality when deciding fair principles by which to live, which I have argued is an inadequate form of justification. This leads Pogge to conclude that Rawls’s global original position cannot claim the full impartiality it supposes. While Pogge suggests that this is because it begins with the justice of states, and not the interests of persons,\textsuperscript{123} I would argue further that it is more fundamentally a result of an attempt to justify via a consensus or agreement. With these considerations we further reinforce the objection that the Rawlsian approach to international justice might exclude some relevant perspectives from issues of justice while failing to claim the necessary objectivity to justify such exclusion.

Moreover, Pogge, like Beitz, finds it becoming increasingly difficult to make a distinction between domestic and international institutions in the way that Rawls’s suggested international original position does. Rather, the deep interconnectedness and interdependence of our global systems and schemes suggest that national boundaries are no longer helpful as the sole category for evaluating the justice of institutions. Our system of states, given the complex nature of relationships between sovereign nations, requires an attempt at global legitimacy.

\textsuperscript{121} Pogge and Labude, “Idea of Justice”, p. 612.
\textsuperscript{122} Pogge and Labude, “Idea of Justice”, p. 612.
\textsuperscript{123} Pogge and Labude, “Idea of Justice”, p. 612.
Like Beitz, Pogge rejects "autarkic" or even partially interactive conceptions of state interaction, arguing instead that states are deeply mutually influencing, and that this influence can potentially be to some individuals' and societies' detriment. He writes, in his article "Human Rights and Human Responsibilities", that "[h]uman rights based responsibilities arise from collaboration in the coercive imposition of any institutional order in which some persons avoidably lack secure access to the objects of their human rights".\footnote{Pogge, "Human Rights and Human Responsibilities", in Andrew Kuper, ed., \textit{Global Rights and Global Responsibilities} (New York: Routledge, 2005), p. 18.} Pogge criticises Rawls's "duty of assistance" on the basis that it entails the problematic implication that actual poorer nations bear the most responsibility for their economic and social problems. This is not only an incorrect assessment of cause and effect in relation to global poverty, but it also leads to the "important moral error" that our only "duty" toward the global poor is one of charity, or that we have only positive duties toward them.\footnote{Thomas Pogge, "Assisting the Global Poor", in Deen K. Chatterjee, ed., \textit{The Ethics of Assistance} (Cambridge: Cambridge University Press, 2004), p. 262.} Instead we are obligated, by our collective complicity in international systems, to aim toward the establishment of the most just institutions we can achieve. Here Pogge belies a crucial assumption contained in Rawls's proposal of a "duty of assistance", that the situation of global poverty can be eradicated, or even ameliorated, without fundamental reform of, or at least comprehensive critique of, the ways in which persons, states and institutions interact across vast distances.

Pogge's critique points to the problematic nature of a conception such as "assistance"; if we can establish that there are other factors involved in the burdening of states besides their own political and economic competencies, as both Pogge and Beitz feel we can, the whole notion then of "assistance" may in fact obscure the history of our current social and economic status quo. Pogge writes that "our enormous economic advantage is deeply tainted by how it accumulated over the course of one historical process that has devastated the societies and cultures of four continents".\footnote{Pogge, "Global Poor", p. 262.} Similarly, the argument suggesting that, because some developing nations have achieved a significant level of growth and

development, those that have not achieved similar results must have internal defects, is, according to Pogge, based on a “some-all fallacy”; the evidence of some is in no way an indication of a potential that is open to all. On the contrary, he adds, “the paths to riches are sparse.”

Pogge’s thesis rests not only on the assertion that the global order contributes to poverty, but also that citizens of the developed world, along with the most well-off citizens of the developing world, not only exist within an unjust global structure but also actively maintain and reinforce such a structure. It is here that his “negative duty” comes in, framed in Kantian terms; if one’s actions are judged to be the cause of harm to another, one is morally obligated to either change one’s action or compensate the victim of such action. If the global system accrues benefit to some while burdening others, those who benefit have a negative duty to work toward either reforming the system toward a more just and fair basic structure, or to compensate (as opposed to “assist”) those in need. Interestingly he observes that this distribution would then not be seen as a “redistribution” from developed to developing societies, but as “offsetting an unjust institutional redistribution from the poor to the rich - re-redistributing, if you like”. While admittedly the entire issue of the link between responsibility and the global structure is still a controversial one, as we will see from O’Neill’s critique, Pogge at least offers strong

---

127 Pogge, “Global Poor”, p. 267. While Pogge acknowledges that in many poor and underdeveloped states corruption and oppression are common practices, he holds that the global institutional order sometimes supports and reinforces these practices. He points to evidence that the current global scheme may in fact (wittingly or unwittingly) discourage ethical and democratic practice in certain political contexts. One way through which the international system might work in this way relates to accepted practices attached to international relations regarding potentially illegitimate possessors of authority. In situations where domestic authority has been sought and maintained through what Pogge terms “the preponderance of the means of coercion”, and where this authority has been recognised internationally, international law permits, firstly, a right to borrow money on behalf of the particular nation in question, and secondly a right to dispose of a nation’s assets without necessarily distributing gains to the citizens. The consequences of such “privileges” can be devastating; unaccountable borrowing and accumulation of short term gains through the disposal of assets can condemn the worst-off of burdened societies to shouldering the responsibility for negligence which brought them no benefit. These observations might lead one to suspect that it is more than inadequate to leave aside the impact of global forces when assessing the problems facing burdened states. While admittedly one can counter this by arguing that the current global scheme does make attempts at sanctioning those states who seem to not be working in the best interests of their citizens (though these sanctioning processes might be criticised for being peculiarly selective), this does not undermine the point that the global scheme contributes in some way to situations of domestic poverty. See Pogge, “Global Poor”, p. 270, and Pogge, “Human Rights”, pp. 19-20.

128 Pogge, “Global Poor”, p. 278.
reasons to suspect that the “duty of assistance” may be an incomplete concept for handling the problem of global poverty.

Finally, the global order as Pogge understands it is not only contributing to issues of global poverty, and not only sustained by developed societies, but, crucially, it is also non-voluntary. He argues that our institutions must be measured according to how well they “fulfil” the requirements of human rights, given that they are often part of a “coercively imposed institutional order”. Under this understanding, when human rights violations occur on an institutional level, we may presume that a negative duty not to harm has been avoided. The duty could then be formulated as “a duty not to contribute to the coercive imposition of any institutional order that avoidably fails human rights, unless one compensates for one’s contribution by working toward reform.”

In conclusion, Pogge suggests three characteristics of the global institutional order: that it is not structured to work in the best interests of the most disadvantaged, that it is reinforced by its members, specifically those who benefit the most, and that it is often non-voluntary, particularly from the perspective of the global poor. These three considerations lead Pogge to conclude that a “duty of assistance” fundamentally misunderstands the nature of global poverty and the appropriate remedies. Instead, he suggests that we include an understanding of collective culpability, something he feels will provide a more adequate starting point for approaches to global justice interesting in overcoming global poverty, or the “becoming well-ordered”, of societies.

Pogge argues that, if we begin an account of international justice from the perspective of a liberal bounded society, we risk overlooking great injustices that may not be perceptible from the perspective of any one domestic society, but visible only as elements of a more fundamental scheme of interaction. This scheme should not be viewed, as Beitz also argues, as one of mutual cooperation; rather,

129 He chooses “fulfil” rights as opposed to “legally enshrining” rights as he argues that legally accessible rights are often practically inaccessible. Pogge, “Human Rights”, pp. 13-14.
participation can take many forms, some of which are voluntary, sustaining, reinforcing and beneficial, others of which are involuntary, forced, compelled and imposed. This complex web of interaction reflects the often-ignored relations between power and vulnerability, something which is either missing from, or incompletely expressed by, the Rawlsian account of international justice.

2.3.3. The inadequacy of the duty of assistance: idealised obligations for idealised societies

If one was to imagine how Rawls would respond to such a critique of the concept of assistance, it may sound something like this; the duty of assistance applies alongside the establishment of a law of peoples whereby all well-ordered societies also establish free and fair economic principles. The global order that has been established is made up of liberal, decent or burdened societies (excluding rogue or despotic regimes), none of whom actively contribute to an unjust scheme. Therefore a negative duty does not apply.

The crucial point in responding to such a defence is to consider where the duty of assistance fits into Rawls's theory of international justice, i.e. how it relates to both ideal and nonideal theory. The duty is introduced as a part of nonideal theory, as a response to "the questions arising from the highly nonideal conditions of our world with its great injustices and widespread social evils" (LP, 89). The role of nonideal theory is transitional; with regard to ideal theory: "[n]onideal theory asks how this long-term goal might be achieved, or worked towards, usually in gradual steps" (LP, 89). Nonideal, or actual world, conditions, do not determine the ideal, rather the ideal is on hand to specify the aims of nonideal theory, which involve transitioning from unfavourable conditions to a full Society of Peoples (cf. LP, 89-90).  

---

What must be noted is that while ideal theory is on hand, ideal circumstances are not. It is fairly uncontroversial to claim that, in order to transition from current actual circumstances to those specified in ideal theory, we must deploy an accurate account of the nature and circumstances of injustice in the nonideal world, i.e. to move forward from injustice, it is crucial to correctly identify its nature and causes.

It is on this point that there appears to be difficulty with Rawls's account. Consider the following comment from Rawls: "[o]n the assumption that there exist in the world some relatively well-ordered peoples, we ask in nonideal theory how these people should act toward non-well-ordered peoples" (LP, 89). It is unclear how we are to interpret this sentence, are we to take "relatively well-ordered peoples" to imply actual states living in the arguably deeply unjust circumstances of the nonideal world, or as well-ordered peoples living under the ideal conditions specified by the hypothetical-ideal establishment of a fair global economic system alongside the law of peoples? Neither option is unproblematic. If it is the former, it must firstly be asked whether any actual states that participate in the current global economic system can be considered well-ordered. Secondly (and more crucially), based on the preceding arguments from Pogge and Beitz, the duty of assistance does not appear to be an adequate response to the implications of nonideal theory, in that it rejects that citizens of liberal societies have duties of justice toward non-domestic agents. These points will be taken up again when we turn to O'Neill's critique of conceptions of assistance. If, however, we take Rawls's comments to imply the latter, well-ordered states living within the context of more ideal global background conditions, then further problems are apparent. I would suggest that there are at least some grounds to suspect that the latter is a more appropriate reading, based on Rawls's later comments regarding the prospects of development for individual societies. Of these prospects he writes that

the causes of wealth of a people and the forms it takes lie in their political culture and in the religious, philosophical, and moral traditions that support the basic structure of their political and social institutions, as well as in the industriousness and cooperative talents of its members, all supported by their political virtues". (LP, 108)
While this comment is intended to respond to critics such as Beitz who have argued for a resource distribution principle, it also, however, has implications for our discussion. Much of the criticism of Rawls’s thinking on the nature of underdevelopment often interprets the above point as the assertion that, if internal factors are the cause, then external factors are not to blame for underdevelopment. It might easily be imagined that the above comment indicates that the two are mutually exclusive, i.e. we either view the causes of underdevelopment as internal, as Rawls apparently does, or reject this and blame external factors, in particular those relating to coercive global forces.

It is, however, unlikely that Rawls believed that powerful global systems, if unjust, could not or would not influence the development of burdened societies. In fact it is clear, from his comments regarding the need to establish a fairer international economic system, that this is not his position. It seems more plausible to assume that Rawls’s comments are contingent on the assumption that a just global order defines, or will define, the background conditions of the burdened societies he refers to. His claims about the causes of underdevelopment would be correct provided we assumed that burdened societies existed alongside ideally well-ordered, developed societies that were embedded in a fair and noncoercive economic order. I would argue that these considerations combined lead to the conclusion that Rawls’s account of burdened societies assumes that they are embedded in an ideal global system that has already established a fair global economic order, i.e. it assumes that certain ideal conditions are already on hand.

There are several issues that arise from this point. Firstly, Rawls’s primary focus appears to be on the establishment of a fair trade regime, with other international issues being handled by what he refers to as “cooperative organisations” (LP, 42). It is plausible to suggest that there are innumerable complex and interrelated issues that affect global systems, and further to suggest that non-state related or trans-state activities, such as those of non-public and borderless financial markets, affect the trade regime in ways that states are, as it stands, powerless to control. While taking account of these considerations is by no means ruled out in Rawls’s theory, it is arguable that their deep impact on the nature and causes of global
justice and injustice demands considerably more attention from an account of international justice than Rawls offers.

Secondly, his strategy for attaining and justifying a fairer system of global trade is questionable. As Pogge and Beitz have argued, it is plausible to assume that at least many among the more vulnerable members of the world’s societies find themselves faced with a coercively imposed and non-voluntary economic system. Rawls suggests that a fairer trade system could be established via a similar strategy to the original position. On the face if it, it seems that it is an unpromising strategy to attempt to justify principles of economic justice via reflective endorsement if the majority of subjects of that system are vulnerable to the extent that conditions for legitimate consent are not met. However, it must be noted that Rawls suggests such a hypothetical scenario only on the condition that “fair background conditions exist.” In this way the system would be non-coercive and voluntary, and presumably any endorsement of the system could therefore (possibly) be considered legitimate. However, rather than rebutting the objection, this claim seems instead to reinforce the primary deficiency of this approach to economic justice. Fair background conditions are ideal predicates, and as we have seen, if Rawls assumes that burdened societies are embedded in the fair global economic system he hypothesises, then these predicates are taken to apply to nonideal circumstances.

This, as we noted, is markedly different to the assumption that ideal theory is already on hand. The assumption of any ideal predicates when considering nonideal theory, i.e. when reasoning about the transitional steps from actual circumstances to more just ones, is problematic. Earlier we reviewed O’Neill’s critique of the use of idealisation in Rawls’s theory of justice. While her comments

133 “Consider fair trade: suppose that liberal people’s assume that, when suitably regulated by a fair background framework, a free competitive-market trading scheme is to everyone’s mutual advantage, at least in the longer run. A further assumption here is that the larger nations with the wealthier economies will not attempt to monopolize the market, or to conspire to form a cartel, or to act as an oligopoly. With these assumptions, and supposing that the veil of ignorance holds... all would agree to fair standards of trade to keep the market free and competitive” (LP, 42-3). Rawls also notes “as in the domestic case... unless fair background conditions exist and are maintained over time from one generation to the next, market transactions will not remain fair, and unjustified inequalities among people will gradually develop. These background conditions and all that they involve have a role analogous to that of the basic structure in domestic society”. (LP, 42)
are in relation to the use of idealisation in the constructivist procedure, they are also relevant to our present discussion. She notes that, while idealisation can be a useful strategy when constructing scientific models, as these will be “tested against the data”, it is a riskier strategy in practical than in theoretical reasoning. Practical reasoning aims to justify norms and standards, to which the world is then to be made to conform: so it cannot be refuted, or checked, by showing that the world does not in fact live up to the proposed norms. Where norms are concerned, we view a level of noncompliance as quite standard, if regrettable, and do not see it as any sort of refutation of a violated norm. Inappropriate idealizations in practical reasoning are not disciplined by the data. Whereas the conclusions of theoretical reasoning must fit the data, the world is to be made to fit the conclusions of practical reasoning.\textsuperscript{134} (Emphasis mine)

Where practical reasoning is concerned with identifying the rights and duties that can guide the transition from nonideal to more ideal circumstances, such a transition will be severely hindered where the identification of rights and duties does not match the actual circumstances in which we find ourselves. The obvious danger is that what actual rights and duties there are will be obscured by the imposition of an ideal depiction of the circumstances of injustice.

A less than just global economic order, which many find themselves with very little option but to “accept”, is a feature of actual nonideal circumstances. The duty of assistance is not an adequate conceptual tool for transitioning from such circumstances to a more just world order, as it is only relevant to idealised burdened societies whose development is primarily hindered by internal deficiencies in their domestic political institutions. This is not the only or even most pressing feature of actual circumstances of injustice. In order for the unjust impact of the actual global system to be reduced, the rights and duties that arise as a result of this system need to be identified. If they are not, there is no way to identify those agents who are responsible for, complicit in, or capable of acting to change an unjust global order. Without the capacity to identify these agents, institutions and processes, there is little hope that progress can be made in reforming the unjust features of the global basic structure: where no one has a

\textsuperscript{134} O’Neill, “The Method of A Theory of Justice”, p. 35.
duty to act, it is not dependable that anyone will. The duty of assistance, rather than correctly identifying morally responsible agents, presents idealised obligations for idealised societies. It is therefore irrelevant to the predicament of actual burdened societies whose circumstances are to a large extent defined by their position in the global economic system. Furthermore, the context in which such a duty may be relevant, i.e. where there exist fair background conditions and a just global economic order, could be described by many as the aim of theories of international justice. We therefore appear to arrive at a paradoxical situation where the aims of justice are assumed as the context of an account of the steps necessary to achieve those aims. The duty of assistance is a wholly inadequate response to the circumstances of injustice, as it applies only to an ideal scenario where many of the salient features of justice have been established.

It is not sufficient to prefix such a duty with the proviso that economic justice has or will be accounted for; a proper analysis of the nature of economic justice (and injustice) highlights certain rights and duties that contradict those established by the duty of assistance. In other words, where an analysis of the actual circumstances of justice appears to yield certain actual duties to reform unjust systems, based on agents’ actions and interactions within the system, the duty of assistance rejects any such causality and instead reduces duty to requirements of assistance or aid, thereby distorting the distinction between justice and benevolence.

2.4. Onora O’Neill on the scope of moral concern and duties of justice

O’Neill also takes up the problem of the relationship between justice and beneficence, which she relates back to the Kantian distinction of perfect and imperfect duty. Like Pogge and Beitz, she argues that there are perfect duties of justice (action necessary to avoid violating principles of justice, as opposed to imperfect duties such as assistance) that extend across borders. However, in order to clarify what is included in such perfect negative duties this section will first address her objections to focusing a theory of international rights or obligations
on complicity or causality in harm (2.4.1). She argues that such an account faces several practical problems in identifying obligation bearers, and may be incomplete if it excludes some who may be entitled to justice, yet whose poverty or need cannot be attributed to harm caused by identifiable agents.

O'Neill suggests a broader and more fundamental approach to the scope of moral concern (2.4.2). While she rejects a focus on harm or complicity in harm, her account shares with Pogge and Beitz the assumption that we owe justice and justification to those whose agency we have a very real possibility of influencing, with whom we share a sphere of interaction. She also notes that an assertion of universal principles alone is ineffective at achieving global justice if there is not also a method for identifying who is a morally relevant other, i.e. for identifying who falls within the relevant sphere of interaction. Her central thesis is that the relevant level of interconnection with others can be identified or revealed in the assumptions we make when planning action; the structure of our action reveals whether we take others to be capable of interacting with us in ways that support or suppress agency, i.e. in ways that make us accountable to them. She concludes that in our contemporary globally integrated context, the structure of our action reveals a "more or less" cosmopolitan scope of moral concern. This also allows us the important insight that if states frequently and persistently fail to meet their obligations, we have good reason to re-consider the system of ascribing all rights and duties to domestic state agencies.

As for whether a cosmopolitan approach to moral concern is compatible with a rights-based approach to justice, she argues that the perspective of rights, i.e. recipience, is less effective than a focus on obligation, i.e. on agency. The demand for universal rights, she argues, will fall on deaf ears if actual and capable obligation-bearers have not also been identified (2.4.3.).

---

135 O'Neill comments that "this relational view points us to a contingently more or less cosmopolitan account of the proper scope of moral concern in some contexts". O'Neill, Bounds of Justice, p. 195.
Turning to the question of what is actually demanded, in our context, by principles of justice, I review O’Neill’s analysis of the principle of rejecting coercion (2.4.4.). She argues that coercive action, by its nature, is easily re-described in more acceptable terms. It is therefore unpromising to attempt to construct an account of coercive action that identifies specific act types as coercive. She argues instead that the more relevant questions relate to the victim’s particular context, i.e. to their particular vulnerabilities or lack of capacities. It is this vulnerability, relative to the power of the potential coercer, that allows for the construction of offers that are perceived as non-refusable. I argue that for this reason it is unhelpful when constructing an account of justice to distinguish sharply between coercive or exploitative action, as both types rely on leveraging the limited capacities of vulnerable agents against them to force (though not necessarily by threat) compliance with an offer that would otherwise be rejected. Based on these considerations O’Neill argues that a principle of rejecting coercion demands that just institutions address the disparities of power that foster opportunities for coercive action.

O’Neill rejects the use of categories such as “aid” or “assistance” to describe the types of duties we bear toward distant others. A focus on aid or assistance obscures or denies that concrete duties of justice exist toward distant others. On a rights-based account such categories are especially problematic because, aside from obscuring duties of justice, such duties of beneficence are seen as “optional”, as they have no counterpart rights. She argues for viewing beneficence from the perspective of an account of the obligations of justice, where it can be seen as selective while also still obligatory, although she emphasises though that even selective (as opposed to optional) beneficence cannot substitute for the demands of justice. She concludes that both justice and beneficence, working within and for reform of the system, are necessary in our contemporary context (2.4.5.).

I treat these aspects of her theory here, as this is necessary in order to highlight the important deficiencies of the duty of assistance: we cannot complete our critique of Rawls’s account of international justice without bringing in O’Neill’s practical arguments regarding justice across and beyond boundaries. While this
section will focus on more practical and applied aspects of O'Neill’s account of justice, Part 3 will analyse the grounding of her account in her distinctive interpretation of Kantian ethics and focus on foundational questions such as the justification of principles of justice.\textsuperscript{136} For now, however, we take for granted that certain obligations do arise from our interactions with others, and address her more practical considerations regarding to whom and in which context such obligations of justice apply and what sorts of actual action is demanded (with our focus being on questions of distributive or material justice).\textsuperscript{137}

2.4.1. A critique of the negative duty not to harm as an incomplete account of obligation

Rawls circumscribed the sphere of morally relevant agents (relevant for the justification of a basic system of rights and duties) within bounded societies. Pogge and Beitz argue instead that our trans-border interactions imply a global basic structure in which we are all complicit and which avoidably and systematically fails to uphold the human rights of many persons. The nature of this structure, it is claimed, demands that we consider distant others as morally relevant in the justification of internal or domestic systems of rights and duties, i.e. that we consider the obligations that arise as a result of our interactions with distant others, which themselves are a result of our engagement with this global structure. As we saw, Pogge points to a number of factors, such as the colonial legacy and the coercive nature of certain international financial practices, to argue for the fundamentally unjust nature of this system.

\textsuperscript{136} In doing so it will aim to highlight the significant advantages of her approach over that of Rawls, in particular how her account of Kantian public reason offers not only a more promising basis for public dialogue in our contemporary world of porous boundaries, but also a more accurate picture of how, why and which kinds of obligations arise from our interactions with others.

Appeal to a negative duty not to harm, to the obligations that arise out of complicity in injustice, harm, exclusion etc., reveal the connection drawn between principles of justice and the scope of moral concern in cosmopolitan accounts such as that of Pogge or Beitz. Principles of justice are universal in scope, and wherever there is a violation of a principle, there is a duty to act. As O’Neill observes, that there are universal principles that extend to all persons is not in general a point of disagreement between communitarian and cosmopolitan positions. Rather, while communitarians do not deny such universality, they also argue that the scope of moral concern is limited to one’s own community, and that this is not incompatible with the idea of universal rights and duties because, they argue, “foreigners will have equal and equally effectively claims against their states, so that accepting their moral standing has no practical implication” (BJ, 187).

Universal rights and duties can be institutionalised within distinct and separate communities without the need for cross-border action on basic principles of justice.

This presents a problem for cosmopolitan accounts that could be expected to argue for the co-extensiveness of moral concern with universal moral principles. Yet universal principles in and of themselves do not get us very far. As O’Neill comments:

[H]ow could we have obligations to millions, indeed billions, of others, and how could they have rights against us? Conversely, could they all have obligations to each of us, and we rights against them? Individually we clearly cannot do much for so many distant strangers, or they for us. If there cannot be obligations to do the impossible, we must seemingly conclude that obligations and rights cannot hold on a global scale between distant strangers. (BJ, 187)

Universal principles are not immediately co-extensive with the scope of moral concern because they themselves do not offer an account of who must act. If it is left vague upon whom distinct duties and rights bear then we cannot simply dismiss communitarians who conclude that if we cannot all act on all duties then we may legitimately circumscribe our action within distinct spheres.

138 O’Neill, Bounds of Justice, p. 187. Further references in text (BJ)
The invocation of a negative duty not to harm seems to simplify the situation somewhat. If it can be demonstrated that “we” are directly or indirectly violating the universally held rights of spatially distant others then it is difficult to disagree with the conclusion that obligations to act (either by refraining or more likely through reform or compensation) arise for us in this context. O’Neill also takes up the idea of rights through complicity, or compensatory justice. She notes that one benefit of such an approach could be that even if the poor lack universal rights to economic support either from fellow citizens or across state boundaries, some of them have special rights to compensation against specified others who have injured or are injuring them. (BJ, 129)

On the point of past injustice, she notes, firstly, that such an account would need a robust account of institutional agency without which there would be no way to identify who holds which rights or obligations, and secondly, even if this could be accomplished, that it “may provide no determinate basis for showing which present individuals, groups, states or regions have special rights to compensation”. With regard to present injustice, she summarises the argument as the claim that the “exercise of political and economic power sometimes helps and sometimes injures the lives of distant and impoverished people in the Third World. Where injury is done, responsibility should lie with the powerful” (BJ, 132). In critiquing this account she argues that while it is “impressive in outline”, its “detailed implications are... obscure” (BJ, 131). She asks what “assumptions about rights... are needed if the operation of economic power is to be seen as rights-violating”, and further, which specific policies do such accounts point towards: reform or compensation? (BJ, 131) Crucially, she asks, “[w]hat is to happen when those who ought to compensate cannot do so?” (BJ, 132) This final point highlights a significant problem: where those obligated to act on this account cannot do so, do

---

139 She also critiques Pogge’s and Beitz’s attempts to reach cosmopolitan aspirations by extending the basic theoretical model deployed by Rawls, and concludes that such accounts suffer from similar deficiencies to the original. See O’Neill, Bounds of justice, pp. 132-4.
140 A seemingly surprising conclusion from these points at the time of the debate in the 1980s, when many countries in Latin America were ruled by military dictatorships, drawn notably by some Latin American economists, is that developing nations should totally detach themselves from the world economy. Among other points, O’Neill argues that detachment will not prevent the global system from seriously influencing possibilities of trade and export. See O. O’Neill, Faces of Hunger (London: Allen & Unwin, 1986), pp. 110-113.
Jill McArdle

those obligations transfer to others, and if so on what grounds, or do they simply fade from view?

It could be objected that the point of present complicity is that we are all responsible and obligated to act, however this seems to return to the initial problem of the indeterminacy of universal principles in relation to action. She concludes that

[to claim special rights we must show a special relationship; but the causal links between specific individuals or institutions who injured or were injured, or who now injure others and suffer injuries, are too often not clear enough to allocate rights of compensation. (BJ, 132)

While her points are persuasive, they do not, I think, serve to entirely dismiss the point made by Pogge: where it is demonstrable that we contribute to systems that cause harm to others we unavoidably enter into a discussion of moral responsibility and obligations to act. However, as she points out, there are serious and complex issues around the controversial implications of such causal relationships for the particular rights and duties of particular persons or groups, and around the interpretation of data; we may find irreducibly conflicting perspectives on whether specific economic or social data point to complicity or absolve “us” from responsibility.

The most compelling point O’Neill raises relates to the incompleteness of an account focused on compensation for past or current injustice. She argues that while many persons are victims of the present, crucially, their poverty and need “cannot be ascribed to any agent, past or present” (BJ, 131). While this point could be addressed again by a cosmopolitan response that the totality of global systems implies all victims are within this system and as such there arise universal obligations, I would argue that it points to how easily arguments grounded in compensatory claims can be undermined where there is an inability to demonstrate causality. Where so much hinges on demonstrating the causal origins of harm, we may struggle to justify more extensive and demanding duties of justice on a global scale. Where causality fails, we may find ourselves still faced with the
claim that we are not committed to seeing distant others as entitled to the same level of justice as those nearer to us and sharing certain domestic institutions. We return to the claim that while principles may be universal, the scope of systems of rights and duties is not. Distant others have rights but "we" need not fulfil them, rather we exist in separate societies that, it might be further argued, can be viewed as distinct spheres of interaction. Crucially, this argument may hold even where it is shown that currently many states fail to meet the needs of their members. This is because, regardless of such failures, the cosmopolitan position has not shown comprehensively that universal principles are co-extensive with the scope of moral concern.

The distinction between universal principles and the scope of moral concern becomes more pressing when we consider that recognition of the negative duty not to harm may be insufficient for identifying those who have actual moral obligations to address the problems of the poor and the needy. Given the incompleteness of such an account, if we are to take the predicament of the world's poor seriously, we may need a broader and more refined account of the relationship between interaction, agency and the scope of moral concern.

2.4.2. The scope of moral concern: the effect of connectivity on the structure of action

While the attempt to justify all cosmopolitan duties of justice through claims of violations of basic universal principles may not be sufficient, the idea underpinning the arguments of Pogge and Beitz may be promising. This idea relates to agency and interaction: patterns of action and interaction bear on agency and in that way have specific implications for setting the scope of moral concern. While their approaches focus in particular on whether we are harming others in our action, O'Neill goes further and looks at the implication of more basic forms of interaction; interaction with others implies a possibility of affecting their agency, and where any such interaction can be identified we are obligated to take the other as a relevant moral agent. For her the scope of moral concern is a
practical question: “what assumptions are we building into our actions, habits, practices and institutions”(BJ, 192). These assumptions are relevant because they reveal the level to which we actually consider other agents capable of interacting with us. In other words, while we may attempt to deny our interconnection with others for the purposes of circumscribing the scope of moral concern and duty, our actions tell a very different story. She summarises the point with the claim that “[i]f in acting we already assume that others are agents and subjects, we can hardly deny this in the next breath” (BJ, 192).

Action reveals how we actually evaluate the capacities of near and distant others, in that

in daily life agents seek to base activity on adequately accurate views about the world and its causal patterns, about their connections to others and about those others’ capacities, capabilities and vulnerabilities, for the solid reason that inaccurate assumptions about any of these may lead to failure, to retaliation or to other harm or injury.\(^{141}\)

In order for us to act effectively in the world, we must at least attempt to make accurate assumptions about the capacities of others who we judge could affect or be affected by our action. Where this is most relevant for accounts of justice is when it is claimed that “our” activities are not sufficiently capable of affecting others to warrant considering them as agents who are owed justice, i.e. when our connection to others is denied. If it can be revealed that there exists connection and possibilities for action and interaction where these are currently denied, there is good reason to think the scope of moral concern must be reconsidered and extended.

There are several ways in which this denial can occur. It can be denied that the other is an agent at all, or held that they are less than an agent, or that while they may be an agent they are at the same time wholly disconnected from one’s sphere of action. These denials are easily revealed as spurious when it is shown that action is shaped by the assumption that the other could act in a multiple of ways, i.e. actual views of the other’s agency are revealed “in the ways in which [those

\(^{141}\) O’Neill, Towards Justice and Virtue, p. 106.
who deny it] organize and adjust their action to take account of others' capacities to act, to suffer and to be influenced" (BJ, 193).

A further form of denial O'Neill discusses reveals not so much how the scope of moral concern can be denied, but how the depth of its implications could be denied. In this case, denial takes the form of "exaggerated views of agency" and is dangerous when it leads to "unreasonable claims about self-sufficiency". That such denials are frequently intended to take advantage of vulnerability and lack of capacity "is revealed by the skill and efficiency with which those who purport to hold these views of others adjust their activities to take advantage of the very limitations they deny". While idealised capacities may be claimed, actual action often reveals an expectation of considerable vulnerability.

On O'Neill’s account, the invocation of state boundaries as a way of justifying a restriction of moral concern is shown to be inconsistent with the actual assumptions that are made about the cross-border interconnectivity of a plurality of agents:

[T]he division of the earth and its inhabitants into mutually exclusive, bounded groups by state boundaries has provided a powerful and systematic way of discounting or at least diminishing the ethical standing of 'foreigners' who live beyond 'our' frontiers, even by those whose activity is in fact premised on the view that those whom they purport to exclude are others to whom they are connected in varied ways. Those who view foreigners as people with whom they can trade, translate and negotiate, reason and remonstrate, whom they can resent and despise, and who can carry complex and intelligent roles, cannot coherently rescind such assumptions of possible connection in order to limit the scope of their ethical consideration, or confine justice within the boundaries of states or communities. (Emphases mine)

In fact she argues that exclusion itself reveals that we assume from others a level of agency relevant to moral concern. Members of more affluent nations assume with regard to others that they

143 O'Neill, Towards Justice and Virtue, p. 110.
144 O'Neill, Towards Justice and Virtue, p. 113.
will not attack or be permitted to settle in their part of the world, and more generally that outsiders will not be permitted to undercut local wages. Put more brutally, a background assumption of most affluent lives is that state power will effectively keep most distant strangers more or less in their place and in their poverty".145

Much of the action that takes place within states, at least in our current historical context, is premised on assumptions that there are many agents beyond the borders of "our" particular society with whom there is a very real possibility of interaction. Furthermore the very act of exclusion shows that "we" view distant others as agents who could interact with us in morally relevant ways, i.e. it reveals an awareness that "our" actions could or can affect them and that their actions could bear on "us".

The level of connectivity revealed by the ways we tailor our action to take account of distant others is not trivial:

[w]e do not and cannot consistently deny the agency of those whose peaceful coexistence, economic sobriety and environmental responsibility we hope to rely upon. (B), 197)

In sum, our actions reveal an assumption that, rather than living in distinct spheres of action, we share a world with others and expect that they could interact with us:

Once we find ourselves, as Kant put it, sharing the surface of the earth with others, we enter into a competition for resources and control, we begin to coordinate and contest with distant others, and we begin to premise our actions, plans and policies on their being agents and subjects. (B), 197)

Our actions make it very clear that we assume distant others to be connected to us via their agency or capacities for action in a great many ways. We owe justice, or justification, to those with whom we share a space for interaction, where there is the very real possibility that we may act on them and they on us in ways that may support or suppress agency. It is not a sufficient rebuttal to this point to suggest a retreat from global integration into nationalist enclaves. This is because the

145 O'Neill, Towards Justice and Virtue, p. 115.
imposition or assertion of a boundary cannot retroactively unravel the scope of moral concern, especially where this action has dire results for the most vulnerable. As we noted, the very act of exclusion indicates that we view or viewed “outsiders” as those with whom we could have interacted. The imposition or assertion of a boundary is then an act whose context is or was very clearly a shared space of possible interaction. Such a boundary must therefore be justified to those others whose moral standing cannot be retroactively denied by attempting to deconstruct the shared sphere of interaction and replace it with more isolated spheres.

However, on this account our connection to others is also not automatically cosmopolitan in scope. It must be emphasised that this is a practical strategy for answering questions in context about the scope of moral concern. By implication then, “[c]onnection will peter out at differing boundaries for different agents and in activities of differing sorts in differing circumstances”. This position is therefore quite distinct from the claim that universal principles automatically imply a universal scope for moral concern. Such a position is ill equipped to answer practical questions about the allocation of duty and responsibility, in particular for questions of global material or distributive justice.

However, O’Neill’s account, while being sufficiently practical and context-sensitive to identify real obligation holders, does imply (given our specific historical context of globally integrated systems) a “more or less cosmopolitan” scope (BJ, 192). It therefore appears to achieve the objectives of moral cosmopolitanism, widening the scope of moral concern, without being vulnerable to the objection that the only practical way to institutionalise moral universalism is to enshrine rights and duties within a system of distinct states. This is where an advantage of O’Neill’s approach is evident. The claim that institutions and duties ought to be cosmopolitan because states frequently fail in their duties is not altogether convincing where moral cosmopolitanism has failed to identify actual obligation-bearers (beyond those who can reliably be identified as contributing to harm).

What O'Neill's account demonstrates is that this type of argument misunderstands the steps of argumentation: it is not argued that duties must be universal *because* states fail; this causality is not implied. Rather, our connection to others, and therefore our relation to them as *bearers of obligations*, is established *before* the response that states as they are can carry the weight of these obligations. The burden of proof therefore lies with *those who claim that the state system is sufficient* to demonstrate that capability. Where it cannot be demonstrated, we revert to the initial position where, given our connection to many and distant others, our obligations have a wide and possibly cosmopolitan scope. If it cannot be shown that the state system is sufficient to meet these obligations, such a system has not been justified and we may legitimately discuss reform or other alternatives in light of our recognised obligations toward each other as moral persons.

2.4.3. Recipience and agency: the rhetoric of rights and the practicality of obligation-centred accounts of justice

O'Neill has argued for a "more or less cosmopolitan" scope for moral concern. She adds to this the claim that such an approach is more suitably accompanied by an account of justice that begins with obligations rather than with rights. Justice is owed to all who fall within the scope of moral concern, however, rights-based accounts argue that justice is best achieved for all via the assertion of universal human rights. While O'Neill at no point rejects or undermines the value of human rights, she suggests that without proper consideration of obligations such reference to rights on their own achieves very little. Her central objection is that

---

147 It is a fundamental aspect of O'Neill's Kantian ethics that obligation, or the question of "[w]hat ought I do?", is the appropriate starting point for a theory of justice. Part 3 will examine her arguments for this particular interpretation of Kant's account of moral principles. Here I only note briefly why she suggests obligation, aside from being the intended starting point for Kant's ethics, is also more *practical* approach to justice.

Practical Reasoning and Transnational Justice

A focus on the perspective of recipience is less effective and practical than a focus on the perspective of the agents who have the capacity to act to fulfil rights:

Because it speaks first to those who claim their rights, and not to those who have the power to meet or spurn their obligations, rights discourse often carries only a vague message to those whose action is needed to secure respect for rights.149

As with the problem of equating universal principles with a universal scope for moral concern, the assertion of universal rights does not tell us who must act to secure them. Working backwards from rights violations to identify the responsible individual or individuals is unlikely to offer practical guidance150:

Since action at a distance is usually institutionally or technologically mediated action, it is hard and often impossible to determine just which individual’s action harms or injures which others, and hence hard to discern who might have rights of redress against whom. (B), 199)

The agent-centred perspective is more practical because it is more directly connected to action:

In beginning with the traditional, Kantian question ‘What ought I (or we) do?’ rather than with the recipients’ question ‘What ought I (or we) get?’ , we face realities more forthrightly, and pose a question that we can address, even if only by beginning the task of constructing institutions. (B), 199)


149 O’Neill, Faces of Hunger, p. 117.

150 It must be noted that this is not Pogge’s position: he uses the claim of complicity in unjust global structure to assert universal obligations, as opposed to suggesting that we attempt to identify and individuate the exact agents responsible for those harms. In this way his position bears more similarity to O’Neill’s in that they both argue that patterns of global interaction greatly widen the scope of moral concern and make most if not all of us morally accountable to distant others. The crucial distinction between their positions is that, where Pogge identifies the relevant interactions as rights-violations, for O’Neill the relevant action is more basic: it is any action that assumes a real possibility of interaction with others whose moral standing is in question. I suspect the difference is trivial in some aspects, meaning that we may infer from Pogge’s claims the point that, where we interact with others, we are morally accountable toward them. Where it does have practical implications is in the evaluation of the demands of universal principles, which on O’Neill’s account demand more than refraining from harm; on her account we do not need to show that the system contributes to harm or that we are complicit to show that we have significant duties to meet the needs of vulnerable agents.
If justice is best achieved through just institutions, the best way forward may not be by merely enshrining universal rights in international charters and tasking inter- or trans-national institutions with reprimanding states who fail to live up to their obligations. Constructing just institutions may be demanding, and if so, an agent-centred starting point that combines principles of justice with cosmopolitan moral scope is in a far better position to consider which concrete steps are necessary to begin constructing just institutions in our world, and who must undertake such steps:

If we can establish some principles of justice, and have at least a practical account of the scope of moral concern, then we may be able to start identifying what is required in order to work towards just institutions... obligations to reduce poverty may be better served by investment, development and educational policies combined with efforts to relieve poverty when economies fail - a mix of institutions may do more than direct attempts to set up ‘rights to food’. (BJ, 199- 200)

Again, this ought not to be read as a denial of such rights, but rather as the convincing assertion that the most effective way to relieve poverty is through an account of justice that is directed at those who have the capacity to relieve poverty. An agent-centred account of principles of justice combined with a practical account of moral concern is the most effective way to begin the seemingly intractable task of addressing global poverty or material justice.

2.4.4. The requirements of justice: preventing coercion, addressing inequality and limiting power

We have so far concluded that the scope of moral concern might be quite extensive, and in our context is more or less cosmopolitan. Taking for granted for the time being that there are principles of justice, I now turn to the question of their demanding implications in the context of actual human conditions: what do the circumstances of justice imply about the requirements of justice?

While O'Neill’s Kantian account of justice argues for several fundamental principles of justice, most notably the rejection of coercion, deception and
violence, I will here focus on the principle of coercion and its implications for our obligations towards others with whom we are connected. O’Neill begins her analysis of coercion by critiquing attempts to identify a necessary and sufficient set of conditions for classifying types of action as coercive. Such efforts are ineffective because coercion by its nature is directed at the specific and particular vulnerabilities of specific and particular agents (BJ, 81-82). The conditions and circumstances that define particular acts or propositions as coercive are heavily dependent on the context of particular situations and as such no uniform set of conditions can be identified or specified in advance. This is particularly important where acts are categorised as either “threats of harm” or “offers of benefit”, because the latter sort can be considered noncoercive and acceptable (BJ, 94). In this way I would suggest that it is important for analysing the difference between coercion and exploitation and the relevance of this distinction for accounts of justice.

Coercion is distinct from violence: some coercion is achieved through violence, however coercion is different in that it must leave some capacity for agency intact in order to secure compliance (which necessarily requires agency) (BJ, 82-3). Violence is also often counter-productive to the aims of coercion, which involve masking the coercive act. It is very difficult to identify and categorise acts as coercive because efficient coercers seek to “redescribe and disown” their actions; act-descriptions can be contrived to offer “acceptable faces” for what may be inherently coercive act-types. The difficulty is further emphasised by the consideration that the success of a coercive act often “hinges on the interpretation of opaque messages” (BJ, 87). Such a feature of coercion, which highlights its subtle and covert nature, undermines the thought that coercive action contains certain features that will be readily identifiable, rather most effective coercive action falls under act-types that are pliant to redescription.

O’Neill suggests instead that when analysing coercion from the perspective of justice
The crucial thing for coercers is the difference between a genuine offer, where choice of one or another offer can be seen as an expression of agency (and can thereby provide the basis for promising, contracting and even for legitimating consent), and a bogus, unrefusable 'offer', where the exercise of choice is corrupted by the offer... A genuine offer, however tempting, however strongly one of its options is preferred to another, can be refused. (B J, 89-90)

The mark of noncoercive interaction is that the capacity of the other to refuse or reject offers is left intact. Their agency is sufficiently unharmed that they can resist the option of compliance. In this way noncoercive action, though it may sometimes seek to affect preference through incentives, does not attempt to intervene in the operation of the will, i.e. to influence actual capacities to resist or reject propositions151 (B J, 89). It is crucial to the legitimacy of consent that the will and capacities to act be left undisturbed: it is well known that the free will is the basis of legitimate consent. Rather than leaving the free will intact by ensuring that rejection of an offer is sustainable for the agent in question, coercers “show profound lack of respect for others’ agency... (and seek) to make residual non-compliant option(s) unsustainable for a particular agent” (B J, 90).

Coercers succeed because they link choice of any but the compliant option to residual option(s) which the particular agents cannot survive or sustain. The coercers’ skill is to identify how to tailor ‘offers’ to the incapacities of particular agents, how to make noncompliant action not merely less preferred but unsustainable, so that their victims are driven to compliance”. (B J, 91. Emphasis mine)

Increased vulnerability and lack of capacity increases the ease with which coercive offers can be constructed. In such situations the variety of options that seem unsustainable may be quite larger than in other circumstances, and there is no guarantee what is considered sustainable to one person will also be perceived as sustainable by an agent in a different context with fewer capacities, “slender resource and multiple commitments” (B J, 92). Based on this consideration O’Neill also argues that

151 Actual refusal does not make coercive acts less coercive, rather “[t]he mark of coercion is an unrefusable ‘offer’, not an unrefusable option”. Even those who show great integrity and refuse coercive offers “have to stay within the framework of the unrefusable ‘offer’”. (B J, 91)
agents who seek not to coerce have to make sure that they do not inadvertently make unrefusable ‘offers’. Any offers they make others must not link options either overtly or covertly to consequences with which those to whom they make the offer cannot live. Like coercers, they will therefore need to take account of others’ strengths and weaknesses, of their specific vulnerabilities and of the actual limits of their capabilities. In particular, they will have to be alert to the ease with which the weak can be coerced. (BJ, 93)

It is with this claim that O'Neill introduces a far more nuanced approach to the concept of coercion. It can be argued that coercion occurs even when the agent who seeks acceptance of an offer does not actually coerce, because it is claimed that the background context is coercive. This claim can be criticised by arguing that unfair background conditions are not sufficient to identify a transaction as coercive, because no harmful action has actually been threatened. Rather than attempt to argue from either of these conventional perspectives, O'Neill has instead refocused the analysis of coercion. She shows that efforts to distinguish coercive action based on criteria such as threats of harm are unproductive, and that the more salient aspect of coercion is the impact on the will through the perceived unsustainability of noncompliance. What her analysis shows is that what is relevant to justice in both cases is that the victim’s agency is radically undermined; they are not realistically in a position to refuse, renegotiate or reject proposals or offers. We then may have good reason to reject accounts of coercion or exploitation that may either consciously or inadvertently question the injustice of certain acts by overly emphasising the distinction between those who threaten and those who “merely” exploit existing vulnerabilities. If the extraction of compliance in each case is dependent on the victim believing that they are not in a position to refuse, then it is a case where agency has not been respected or protected and it is clearly a matter of a violation of fundamental moral principles.

It may seem that exploitation is less proactive than coercive action, however we have seen that coercers do not construct or create the vulnerabilities that are so crucial to their purposes. Rather, coercers frequently identify and exploit specific, pre-existing vulnerabilities by utilising them as leverage devices in constructing offers. Coercers exploit and take advantage of vulnerability; the argument that only
those who take the further step of threatening to actually undertake certain harmful actions qualify as violating basic principles of justice is unconvincing. In other words:

If the context and content of individual decision-making is determined by social and economic structures which even sometimes rely on others’ need in order to secure ‘consent’ or ‘agreement’, these arrangements may be fundamentally coercive and deceptive.152 (Emphasis mine)

It may suit those who would argue for the sufficiency of much current commercial bargaining practice to focus on the identification of actual threats of harm, but from the point of view of taking principles of justice seriously we can see that such a distinction does not move us very far toward understanding the salient aspects of coercion or exploitation. O’Neill has shown that adherence to standard forms of commercial bargaining may be “fundamentally coercive” (FH, 147). Where vulnerability is great, i.e. in the actual circumstances of human life, such practices are “a skimpy and formalistic substitute for fundamentally noncoercive and nondeceptive forms of life” (FH, 149).

O’Neill also makes the important point that vulnerability is “relative”; victims are vulnerable in that “they possess fewer capabilities, powers or resources than others, and specifically than their coercers” (BJ, 95). This implies that differentials of power are crucial to thinking about justice, because

[the prospects for making unrefusable ‘offers’ will always be numerous and varied, and more numerous and more varied where differentials of power are greater” (BJ, 96)

Where persons or agents are deeply vulnerable in relation to those with whom they interact, through differentials of power, there is endless opportunity for coercion and exploitation, i.e. to achieve ends by relying on others’ specific weaknesses to limit their capacity to refuse. This implies for O’Neill that while a strategy of limiting misuses of power through law, regulation, equal rights and democratic governance is important, it makes sense to simultaneously attempt to reduce the differentials in power themselves (BJ, 96). I would suggest that such a strategy is

152 O’Neill, Faces of Hunger, p. 148. Further references in the text (FH)
Practical Reasoning and Transnational Justice

required in a world where law and regulation can barely identify let alone eradicate all coercive acts, practices and systems, many of whom can be easily redescribed as acceptable and legal dealings and transactions.

On this basis it can be argued that material and distributive justice are critical concerns for the realisation of principles of justice, i.e. for realising our obligations towards others and constructing just institutions. Moving closer to justice means taking "account of the varied ways in which human agents whose needs are unmet are vulnerable to forms of deception and coercion" (FH, 149). The distribution of rights and goods to meet the needs of the most vulnerable persons is therefore a matter of justice:

Any just global order must at least meet standards of material justice and provide for the basic material needs in whose absence all human beings are overwhelmingly vulnerable to coercion and deception. (FH, 141)

Without just institutions designed to meet the needs of vulnerable persons we will make little progress on issues of injustice, as regulation alone cannot eradicate the vulnerability that makes coercive action inevitable:

Justice is embodied in public institutions and policies which secure freedom from deep forms of coercion and deception. Circumstances of justice are lacking so long as material and social needs are so great that coercion and deception are not merely easy but virtually unavoidable". (FH, 146. Emphasis mine)

What all this implies for our discussion of our duties and obligations to distant others is that, as was argued by Pogge and Beitz, the duty of assistance is not the appropriate category for understanding our obligations to distant others. Meeting the needs of distant others is not a matter of "aid" and such categorisations undermine and divert attention from the actual obligations we hold as a matter of justice: when obligations are parsed in the language of aid and assistance "it is implied that they are not required by justice" (FH, 151. Emphasis mine).
2.4.5. The priority of justice over beneficence: assistance and obligation

The substitution of beneficence for justice is problematic on any account, but it is especially problematic on rights-based accounts. Beneficence falls under the category of imperfect duties, meaning its enactment cannot be demanded by specific others. There is no claimable right that serves as a counterpart for imperfect obligations. What this implies is that, from a rights-based perspective, imperfect duties offer little... Their performance can neither be claimed nor waived, and is not required by justice but is a matter of charity or optional beneficence. (FH, 102)

Rights-based accounts cannot ground as obligatory anything that does not derive from the perspective of recipient entitlement. Obligations stem from what others can legitimately claim: charitable action cannot be claimed by others, there is therefore no basis on which to view it as obligatory. Beneficence is then unlikely to get us very far in transitioning toward more just circumstances and institutions, and simply describing aid or assistance as a “duty” does not override the fundamentally non-obligatory nature of such a duty on rights-based accounts. The only other way forward on such accounts is to claim human rights to have basic needs met (such as the right to food), however, as we have seen, these do not get very far either if we cannot identify capable obligation bearers (keeping in mind that allocating obligations to states is often ineffective for practical purposes). In such circumstances even extreme poverty will only be a matter of injustice “if [victims] have special rights to have their material needs met”, and as we have seen, this is not a promising approach to handling issues such as global poverty (FH, 102).

It appears that if we begin from a rights-based position, linking material need and poverty to those who can and should act as a matter of justice is problematic from the perspective of both justice and beneficence. Issues such as world hunger are either a matter of justice for which there is a claimable right but no effective

---

195 Whereas perfect duties can be demanded by others: “On the Kantian account perfect duty is action which is required in a given context in order to avoid acting on a non-universalizable fundamental principle; most such duties are duties of justice”. (FH, 138)
obligation linking those in need with those who could act, or it is a matter of beneficence, in which case action on world hunger is entirely optional. In comparison to rights, imperfect duties such as "beneficence and help are likely to seem less important, especially in public affairs" (FH, 102). This means that if we cannot ground an effective global right to have basic needs met, then matters of poverty and hunger are demoted from the sphere of justice into a sphere of non-obligatory optionality. In this case the likelihood of ever achieving more just circumstances is radically undermined.

O'Neill's account, which begins from principles of justice and obligation, offers a different understanding of the nature of imperfect duties.\textsuperscript{154} She argues that imperfect obligations must still be viewed as obligatory, because without imperfect obligations such as help and respect "the agency of limited and needy beings is insecure" (FH, 141). We cannot hope to respect and secure agency where imperfect duties are entirely eschewed. Agency is always vulnerable without an expectation of a certain amount of help from others, and this counts even in circumstances of justice (FH, 141):

the practice of just institutions may fall short of their principles and.. even those who live in circumstances of justice are often unable to act without the help of others. (FH, 141)

Imperfect duties are obligatory, however, they differ from perfect duties in that their performance cannot be required in all circumstances to which they may apply. In the case of help for example, it is not possible to provide assistance at all times to all those who may need it. The fallacy that arises from rights-based accounts is that this limitation implies optionality. O'Neill points out that what is actually implied is selectivity. We are not obligated to fulfil duties of help in all possible contexts, however, this does not imply the option to reject such duties entirely. Full omission of imperfect duties is wrong (FH, 145); they are not optional extras although they are "unavoidably selective" (FH, 146. Emphasis mine).

\textsuperscript{154} For a more comprehensive account of the nature of perfect and imperfect duties, see O. O'Neill, \textit{Acting on Principle: an Essay on Kantian Ethics} (Cambridge: Cambridge University Press, 2013).
It is also a fallacy to assume that because beneficence is obligatory, it can fulfil the role of justice. As we have seen, beneficence is not justice and there are obligations of justice toward distant and vulnerable others. Beneficence is a weak and ineffective substitute for the demands of justice, as imperfect duties cannot “provide the institutional conditions which systematically meet material needs and guarantee the absence of coercion and deception” (FH, 146). Beneficence works within the system while justice challenges the system and the practices it institutionalises:

Justice is the more fundamental part of obligation because it concerns the framework of institutions and practices which form the context of actions and make certain problems salient, certain solutions possible and certain modes of thought available. (FH, 161)

It is where supposedly just systems and efforts to reform such systems fail that beneficence and aid become more demanding obligations; in a less than just world imperfect duties are more urgent than they might otherwise be (FH, 159-60). In this way we can say that while a (necessarily selective) performance of imperfect duties is required in contexts of injustice, it cannot and must not take the place of actual justice (FH, 160). In practical terms this implies that we cannot involve ourselves in aid or assistance while supporting “a world political and economic system which is an obstacle to justice” (FH, 162). Beneficence (within the system) and political activity (to change the system) are not incompatible; rather a commitment to charitable action is also a commitment to material justice, which implies a commitment to political change (FH, 163).

The deficiencies of the duty of assistance are now evident. On Rawls’s account the entirety of “our” obligations to non-domestic persons is encompassed by this apparent “duty” to assist “burdened” or underdeveloped societies. This duty is at best an imperfect duty of beneficence. As the whole of transnational obligation, it therefore denies that we have any obligations of justice that extend beyond our own borders (other than perhaps a duty of fair dealing in trade affairs). As we have seen, there are two pressing problems with this account. The first is that beneficence on a rights-based account can be seen as merely optional. Granted,
Rawls frames beneficence as a duty that must be carried out wherever societies are burdened by unfavourable conditions, however, he offers no reasons for viewing it as such. There is nothing in The Law of Peoples that grounds the obligatory nature of assistance or aid. Without any account of why duties of help or assistance are obligatory, i.e. why we should take the agency of distant others seriously, such a duty can be easily dismissed as optional.

The second objection is that beneficence, as the whole of international obligation, substitutes the role of justice. By asserting a duty of assistance Rawls effectively denies that there exist any more substantial duties of justice toward distant others. While he does not deny that such others are moral persons, he follows the communitarian reasoning that each state can legitimately circumscribe a particular set of agents with a system of rights and duties. Therefore, while distant others may be entitled to claim certain rights, they may not claim them from “us”.

Pogge and Beitz demonstrate clearly that the patterns of interactions that define much of the global structure potentially make us accountable to many beyond our domestic borders, because we interact with them, sometimes through institutions, in ways that can have deep and lasting effects on their agency. O'Neill deepens and reinforces these considerations by highlighting how our actions reveal the assumptions we make about others’ capacities to act and react, i.e. about their agency, and how we cannot consistently deny relevant interaction while presupposing it in our action. Together they provide a convincing account of the more or less cosmopolitan scope of moral concern: we owe justice to those with whom we share a sphere of interaction where there are very real possibilities for influencing others’ agency.

With the presupposition that we have strong reasons to consider distant strangers as morally relevant others, O'Neill argues persuasively that beneficence is not

---

155 Rawls comments on the motivation for the duty of assistance: “[w]e take as a basic characteristic of well-ordered peoples that they wish to live in a world in which all peoples accept and follow the (ideal of the) Law of Peoples” (LP, 89). This is not a particularly convincing account of the obligatoriness of duties toward distant others; rather it seems to undermine justice by predicking it on the wishes of liberal peoples for reciprocal support of an international framework, rather than a striving to aim for justice also at a global level.
optional but selectively obligatory, and that furthermore, it is a weak substitute for
the very real obligations of justice we bear toward members of other societies.
Beneficence only becomes the centre of actual obligation where just institutions
fail, and in this context it must be seen as obligatory and demanding.

O’Neill also demonstrates how issues such as global poverty and material need are
issues of justice and not charity. Her account of coercion argues that the key
feature of coercive contexts is the relative vulnerability of the victim to those who
either seek to threaten or exploit them. An effective account of justice must
therefore consider how differentials of power can be reduced. Meeting material
needs can equalise power relations and reduce the forms of vulnerability that
make coercive acts and practices all but inevitable.

National boundaries do not define the limits of our obligations. We are connected
across vast distances with many others and where there are possibilities for action
and interaction, there are possibilities for either supporting or undermining
agency. In this context the claim that justice is the responsibility of states is deeply
inadequate. This does not amount to the claim that boundaries are always
necessarily unjust. However while they are not necessarily so, in the actual context
in which we find ourselves they frequently institutionalise forms of exclusion that
have a detrimental impact on agency (cf. BJ, 200):

The exclusions boundaries inflict are surely not inevitably unjust, in that all those whose standing I
am committed to acknowledge might in theory have protections and prospects in their states
which meet the demands of justice. On a rosy view, this is what the system of states is to achieve.
Yet we know that it wholly fails to do so. Demands for intervention in cases of violations of basic
rights, demands for secession, demands for asylum show that for many millions, current inclusions
and exclusions are seen as sources of deep injury. (BJ, 200)

Boundaries are not inherently unjust, however, we cannot conclude from this that
we may assume all exclusions are justified in advance; rather we must treat
“outsiders” as persons who are owed justification. If we combine principles of
justice with the appropriate scope of moral concern we have a starting point for
Practical Reasoning and Transnational Justice

building just institutions. This does not mean dismantling boundaries, rather it implies

considering carefully to whom and to what (to movements of persons, of goods, of information, of money) any given boundary should be porous. Porosity is endlessly variable and adjustable; different filters can be institutionalized. (BJ, 200)

Recognising that we hold demanding obligations to distant others does not by itself supply a necessary set of institutional reforms, rather we must consider the many possible ways which we can effectively attempt to meet obligations, which may include either making boundaries "more porous in specific ways or by compensating [others] for any harms caused by otherwise unjustifiable exclusion" (BJ, 202). Whatever avenue is chosen, what is crucial is the recognition that exclusion via boundaries is not a presupposition of justice and must be justified to those who are excluded. Correspondingly, it must be recognised that transnational institutions (both actual and possible) tasked with institutionalising principles of justice, fulfilling human rights and meeting the needs of vulnerable persons across the world are grounded in non-optional duties of justice, and not in weaker conceptions such as "aid" or "assistance".

2.5. Conclusion

Having argued that the principles of justice make heavy demands in the context of connected agents who possess varying and disparate levels of capacities (and incapacities), it remains to be demonstrated how any principles of justice might be justified. Rawls argued that the perspective of reasonable persons was the most objective perspective available to ground justification. Despite all the considerations of the past section, it may still seem that Rawls’s approach to objectivity is the only available option, and to claim that it is not properly objective is merely to compare it to a perspective that is unavailable, for example to compare it to Kant’s claim, rejected by Rawls, that reason can supply its own conditions of justification. I would suggest that this response rests on the
assumption that, despite not being *fully* objective, Rawls's approach is at least a practical and workable approach to reasoning about justice. With this in mind I review the conclusions of our critique.

I have argued that as a result of Rawls's justificatory strategy, his account of justice has no defence against those who would claim that it cannot provide a justifiably objective perspective. As Rawls himself noted, agreement, consent or consensus cannot be presumed to justify anything if the conditions for their legitimacy have not been met. Rawls attempts to *supply* such conditions, but ultimately begs the question by *justifying them via an actual agreement*. It is his refusal to consider that something *other than consent* is fundamental to justification that leads him to this curious justificatory impasse, where he acknowledges that actual consent is not fundamental yet still justifies the conditions for hypothetical consent via an actual agreement.

I briefly reviewed O'Neill's account of the justification of political institutions in order to demonstrate that an appeal to consent, whether actual or hypothetical, was not necessary to justify basic principles such as freedom and equality in a constitutional state. Rather such principles were the *necessary conditions for all legitimate agreement and consent*. Despite recognising that certain conditions were necessary to legitimate consent, Rawls's focus on consent itself (both actual and hypothetical) means that he ultimately leverages or hinges the validity of the conditions for consent on what *certain persons actually agree to*. In this way he undermines any claim that certain principles, duties or conditions are necessary. As demonstrated by O'Neill, however, there are grounds to think that certain principles are more fundamental than considerations of consent. Their justification is more fundamental than a consideration of what rational and reasonable persons would agree to; they form the account of conditions without which *no agreement can be presumed to have any validity*.

It was also argued that these were no mere theoretical considerations. That practical and public reason cannot be justified *outside the consensus by which they are defined* has considerable practical import. Taking the case of those excluded on
the basis that they are "unreasonable", I argued that Rawls's account failed to supply the necessary independent objective perspective on which to ground such a claim. Appealing to what "we" agree to in order to justify the exclusion of those who disagree is no form of justification. Regardless of how unreasonable those excluded may be, if objectivity and legitimacy are confined to the consensus it is the prerogative of those outside to declare the agreement unjustified. It may seem irrelevant to "us" whether those outside declare the consensus unjustified, if "we" are convinced of our own reasonableness. However I would suggest that in an increasingly integrated and radically pluralist world, such an approach is not promising.

From our critique of the account of decency, it was concluded that Rawls's account leads to a refusal to critique institutions that clearly stand in need of criticism, given that they reject the conditions of legitimacy. The basis for this acceptance of what might otherwise be declared "unreasonable" was the notion of a "people" and its sovereign integrity. I argued that without meeting the basic conditions of freedom and equality the notion of a "people" as a collective unit sharing in a conception of justice was not a valid assumption. I would add now that such a notion, if unaccompanied by universal-moral justification and an account of the conditions for legitimately shared agreement, is a dangerous form of rhetoric. Rawls's strategy of justification, by forsaking universal validity, is a gift to those who would seek to reject principles of freedom and equality and to mask injustice by describing it as a legitimate alternative to an ethnocentrically Western system of thought. As to the point that we may exclude those others as unreasonable, I would reiterate my concern about the impracticality of such a strategy. Retreating from critical engagement with the other is not an option where we share a sphere for possible interaction. It is not possible to dismiss others whom we cannot avoid influencing, indeed it is a familiar reality of our contemporary world that many whose accounts of justice "we" may abhor are also those whom we are forced to rely on for trade and security purposes. Neither retreat nor acquiescence seem possible or preferable, rather critical engagement seems the only way to sustain communication without forsaking justice. We cannot simply avoid those who disagree with us, and liberal institutions and basic human rights must be
Justifiable to those who would reject them. If they are not, then the claim that they are applicable and relevant to all moral persons is radically undermined.

As for the exclusion of others on the basis that they are not located within the boundary that is (apparently) contiguous with the consensus, such exclusion, without independent, universally accessible justification, amounts to relying on the consensus to justify itself. To define the limits of the bounded society by reference to what those inside would accept, endorse or affirm (by appealing to shared agreement to ground the political institutions of that society) is to take consensus as fundamental to its own justification. The boundary marks the limits of the consensus, and if the consensus is not fundamental to justification then neither can its limits be taken for granted. The very act of exclusion implies that the limits could have been drawn differently, and therefore this particular state of affairs, this particular set of boundaries, must be justified independently of the fact of its existence. What is needed is a form of justification that does not rely on the perspective of those who seek to exclude.

Such a claim is of course predicated on the assumption that those others are relevant with regard to the scope of moral concern. Our analysis of this scope highlighted that, in a world of complex global interaction, the domestic political institutions of bounded societies may not be the sole or even primary influences on the capacities, freedoms and opportunities that theories of justice seek to protect and support. Rather, agency is affected by a wide and complex variety of interactions, institutions and global networks, many of which transcend the institutional powers of the bounded society. A practical approach to the scope of moral concern must attempt to identify the contexts where persons interact with each other in ways that can undermine or promote agency. As we saw, such connectivity can be identified by assessments of action; all action must make accurate assumptions about others' capacities in order to succeed, and it is these assumptions that reveal our level of interconnectivity with others. It is frequently evident from the ways individuals, organisations and institutions structure their action that they expect the possibility of interaction with a wide range of others, many of whom may be spatially separated from them by great distances. In this
way the scope of moral concern is quite wide and in some contexts relatively cosmopolitan.

In sum, we cannot do without a more objective and universal account of reason for the justification of principles of justice, and this objectivity has demanding implications for the realisation of justice. O'Neill argues that Rawls's account of reason does not qualify as public reason at all, and cannot ground any real principles of justice. Before turning to her account of public reason we will address her interpretation of Kant's account of practical reason and principles of justice. Like Rawls, she is concerned with Kant's account of reason, freedom and autonomy. However, she rejects entirely the attempt to detach Kant's account from the claim that reason itself can supply the conditions for its own possibility. Instead, she argues that this claim is the foundation of Kant's constructivist account of justice. If she can vindicate such a position, her account could provide the non-partial and universally justified perspective that Rawls's consensus-based strategy claimed was unavailable.
Part 3

Onora O’Neill’s framework of Kantian cosmopolitanism

Part 2 argued that strategies of justification that deploy contingent premises are incapable of providing a universally accessible foundation for the principles and duties of justice, and hence can offer no basis for *reasoning* about justice across traditions and cultures. Part 3 attempts to elucidate O’Neill’s alternative proposal for understanding Kant’s justification of principles of justice via his account of practical reasoning. This analysis proceeds in two parts. Firstly, I review her particular interpretation of Kant’s justification of principles of justice by examining his fundamental conceptions of freedom, reason, autonomy and agency, and how they relate to the different formulations the Categorical Imperative (3.1.) I discuss O’Neill’s claim that Kant’s account of justice must be interpreted as politically orientated, though political must be understood to relate to the abstract social context within which all persons find themselves, and not, for example, to a specific liberal democratic society. For O’Neill, the Kantian conception of freedom is a practical while not empirical conception that must be presupposed by both theoretical and practical moral reasoning. Practical reason itself is interpreted as the capacity of reason to legislate itself, autonomously, according to practical considerations about the possibility or impossibility of a plurality of *agents* acting on certain principles. Several of the formulations of the Categorical Imperative are then to be understood as equivalent expressions of these underlying conditions. Kant’s justification for principles of justice, i.e. the criterion of universalizability, is not grounded in considerations for seeking hypothetical or actual agreement. Rather, this account emphasises what *could* be agreed to by a plurality of agents by arguing for the universally obligatory nature of principles that respect human agency.
Having reviewed the foundations of O'Neill’s Kantian interpretation, I turn to address the question of public reason (3.2.). Rawls’s account of public reason is first examined, and rejected on the basis that it cannot provide an accessible foundation for practical reasoning in the context of real ethical plurality, where persons do not already share certain premises. I question whether an assumption of cultural relativism is a convincing rebuttal of this criticism, and conclude that given our actual historical context relativism is neither convincing nor promising. Finally, O'Neill’s understanding of Kant’s account of public reason is addressed. In particular this analysis aims to draw attention to the fact that, on her account, principles of justice, public reasons and maxims of communications are justified by the same guiding concerns: that reasons must, in principle, be followable and shareable by others. Reasons that cannot be shared or accepted by others, either by appealing to some contingent external authority or by proposing action that undercuts agency, are not suitable as public reasons. The authority of reason can only be established in a reflexive process of examining reasons (both as an individual bearer of practical reason and in discourse with others), for their capacity to be shared by a plurality. Public reasons, unlike private reasons, cannot then restrict themselves by appeal to shared premises; rather they must aim at the possibility of acceptance by the widest audience.

3.1. O'Neill’s interpretation of the foundations of Kant’s ethics

To understand the particular interpretation of Kantian ethics adopted by O'Neill, it is helpful to begin by noting the ethical and philosophical positions she herself identifies as opposite to hers; approaches which “treat rights rather than obligations as fundamental... and rely on a preference-based theory of action and an instrumental account of rationality”, and which also deny the possibility of discovering universal principles.\footnote{Onora O'Neill, “Kantian Ethics”, in Peter Singer (ed.), A Companion to Ethics (Oxford: Blackwell 1993), pp. 175-184, p. 184.} The approach O'Neill wishes to set forth as an alternative then is one which has the resources to incorporate both individual capacities for moral decision making and institutions of justice, by taking...
obligation as its starting point and having as its foundation a theory of action and rationality that relies upon substantial accounts of human freedom and autonomy. Such an approach, she argues, will yield a reliable method for determining principles by which a plurality of “interacting agents” can live together.\textsuperscript{157}

This chapter will begin with a brief look at O’Neill’s assertion that the use of political metaphor in Kant’s writing is central to his vindication of practical reason, because it is the task of constructing principles for a plurality of agents that provides the justification for the authority of practical reason (3.1.1.). I will then address the family of accounts of instrumental reasoning and subjective agency that O’Neill investigates and subsequently rejects in favour of a non-instrumental account of practical reason that links agency with moral decision-making (3.1.2.). I will then turn to her analysis of Kant’s defence of human freedom (3.1.3.). This section will address how Kant defends freedom on practical (rather than theoretical) grounds, and the division O’Neill identifies between positive and negative freedom in Kant’s account. This will lead then to O’Neill’s interpretation of the formulations of the Categorical Imperative, as it is her assertion that several of these are equivalent formulations of the realisation of \textit{positive} human freedom (3.1.4.). I will conclude with some reflections on the advantages and implications of O’Neill’s reinterpreted Kantian framework (3.1.5.).

\textbf{3.2.1 The significance of the “political metaphor” in Kant’s vindication of the authority of practical reason}

What is significant in O’Neill’s particular Kantian approach is its emphasis on the practical, political dimension of Kant’s ethics. While some have interpreted the Categorical Imperative as a solipsist reflection, O’Neill is instead arguing that his approach is intrinsically, and fundamentally, social. This is because, according to O’Neill, he \textit{begins} with what she calls “the political metaphor”, a term that stands

for the image of a plurality of agents attempting to construct principles by which they can live together.¹⁵⁸ In Constructions of Reason, O'Neill sets out to differentiate Kant's methodological starting point from what she identifies as rationalist, or Cartesian, foundations. Specifically she argues that Kant does not follow Descartes’s individualist or autobiographical starting point, rather she suggests that Kant avoids such a position by beginning with a practical, political task (CR, 5). According to O'Neill, he achieves this through the use of political metaphor. She asserts that "[t]he central point that Kant makes with these analogies is that reason's authority must (since it receives no antecedent or transcendent vindication) be seen as a practical and collective task", like that of constituting political authority" (CR, 18. emphasis mine).

The “political metaphor” points us toward what will be the vindication of the rule-giving authority of reason, by framing Kant’s approach in practical and political terms. She suggests that “all that we get initially is a gesture toward the thought that a critique of reason must be a reflexive and political task” (CR, 9). Further on, she summarises his use of political images in the final part of the Critique of Pure Reason as follows:

When Kant subverts the Cartesian images of construction for his own purposes at the beginning of the “Transcendental Doctrine of Method”, he suggests that the “building” of human knowledge can use only available “materials” and must follow “a plan” that is not antecedently given, but must be devised and deployed by a plurality of agents who share a world, but who are short of principles for doing the sharing. This is why the basic task of constructing principles of discursive order is analogous to that of constructing principles of political order, and why politics provides metaphors for articulating the task, principles and limits of reason. (CR, 20. Emphasis mine)

According to O'Neill then, Kant sees his account of the construction of principles of justice as very much a practical “task” or project of construction, over against an effort that remains only at the level of theory (CR, 7). This practical perspective provides the justification for the content of Kant’s account of morality, and there are two ways in which this is particularly significant. Firstly, in relation to his

account of freedom, it is this practical standpoint that provides the justification for his defence of freedom; as I will briefly summarise, it is for practical purposes that we may conceive of ourselves as free.

Secondly, in relation to the Categorical Imperative, O'Neill points out how each formulation relies on this politically orientated starting point for its justification. Of the agents in question, those who are to devise a plan for the construction of shared principles, O'Neill notes that

none of them can (if they are to arrive at any plan and to build even a modest structure) act in ways that will rule out arriving at a plan. If they are to have any sort of collaboration and construct any sort of 'building', they must at least not act in ways that undermine the possibility of collaboration. (CR, 21)

O'Neill also comments that “the constraints on possibilities of construction are imposed by the fact that the principles are to be found for a plurality of possible voices or agents who share a world” (CR, 27. Emphasis mine); it is in asking why we seek shared principles then that we discover what criteria will apply to those principles.

What I will endeavour to illustrate in the following chapter is O'Neill’s interpretation of the way in which Kant’s accounts of agency, autonomy, freedom and reason interact to provide us with a framework for coordinating action amongst a plurality of agents, about whom we make no assumptions beyond some basic general conditions of humanity, i.e. interdependence, rationality, finitude and vulnerability.159 As we shall see, the accounts offered are inextricably intertwined; each is equally reliant upon the rest. None can be understood in isolation; rather they must be understood as compatible and mutually contingent expressions of the basic principles of Kantian morality. Before introducing the Kantian approach adopted by O'Neill I will briefly address some of the positions O'Neill rejects, as these provide indications of the strengths of her Kantian alternative (3.1.2.). I will then turn to Kant’s understanding of the notion of freedom (3.1.3.), which will

---

introduce the link between freedom and autonomy, thereby leading us to his formulations of the Categorical Imperative (3.1.4.). This will allow us to conclude with a summary of the interrelationship of his fundamental conceptions of agency, autonomy, freedom and reason, and the implications this carries for O'Neill's particular approach to justice, in particular her criticism of the approach to public reason set forth by Rawls.

3.1.2. Instrumental reasoning and empirical accounts of agency

O'Neill situates her particular form of Kantianism, which takes the shape of action-based reasoning, in opposition to a family of accounts of practical reason and agency, all of whom she rejects as inadequate or incomplete. In what follows I will briefly summarise the accounts she rejects, and on what grounds she rejects them. The criteria she cites as lacking in these approaches will help to highlight what she finds so advantageous in her interpretation of the Kantian approach to reasoning about ethics. The two primary conclusions she draws are, firstly, that results-based reasoning is indeterminate and arbitrary in relation to the worth of differing ends, and, secondly, that the linking of preference to action as cause to effect is inadequate as an account of agency.

O'Neill rejects teleological or end-orientated conceptions of practical reason. She divides them into two kinds. The first conception claims to be directed toward some objective ends, and she describes this as belonging to the Platonic tradition (cf. BJ, 18). Such objective ends, according to O'Neill, invoke an "arbitrary and illusory authority", and as such they are simply "unconvincing" as justifications for action (BJ, 13). Secondly she addresses what could be considered a less controversial teleological account of practical reason in modernity, which claims that there are no objective ends, only subjective ends, and that the motivation of practical reason is to determine action for achieving whatever subjective ends.

160 This will be a less than exhaustive account, as the aim here is simply to give a brief account of the positions rejected by O'Neill in order to frame her adoption of Kantian universalism. For further reading please see O'Neill, Bounds of Justice, Ch. 4. "Kant's justice and Kantian justice", pp. 65-80, and O'Neill, Faces of Hunger, Ch. 4. "Reasoning about Results", pp. 52-70.
there are. On this account we find that “justification will attach to means, and motivation will flow from subjective ends” (BJ, 17). The appeal is evident; such an approach removes difficulties relating to defining any objective good, and limits itself to the mapping of means to ends. There are two major objections O’Neill makes to this approach, one relating to means and the other relating to ends. I will begin with the former, her rejection of result-orientated reasoning, before turning to the latter, her rejection of the preference-based theories of action that are implied by a subjective teleological account of practical reason. I will address the deficiencies of result-orientated reasoning, specifically in relation to the identification of categories of action and their justification (3.1.2.1). I will then treat her critique of preference-based theories of action, focusing on what she sees as the inability of such theories to identify the meaning or worth of specific acts in relation to others (3.1.2.2). Lastly, I will address the role of autonomy in preference-based theories of action (3.1.2.3.). O’Neill argues that such theories fail to offer any account of autonomy that can explain why autonomous action is valuable in relation to morality or justice.

3.1.2.1. Result-orientated reasoning as a deficient form of practical reasoning

O’Neill claims that for instrumental accounts of practical reasoning to be enacted or applied to real situations, it is typical for them to “begin from some listing of the ‘available’ options” (BJ, 18), which, as she puts it, generally “incorporate the socially accepted and prized categories of action which participants view as the ‘real’ options for a given situation” (BJ, 18). What is of importance here is that these categories are often unvindicated, preference-based categories. In her book on applied ethics, Faces of Hunger, O’Neill describes how such result-orientated models are dependent upon the pre-existence of particular act-descriptions or norms that can narrow the possibilities for determining action. For action to be determined through instrumental result-centred reasoning, the set of available actions must first be known. O’Neill suggests that what tends to occur in practice is

161 O’Neill, Faces of Hunger, Ch.4.
that, because we cannot identify the fullest possible list of acts, restricted lists of act-descriptions are instead adopted by result-centred reasoners from other sources, often from "practical principles and social norms" (BJ, 18). So while such reasoning may be able to provide general agreement on the identification of broad problems that require resolutions, the available actions that could be prescribed tend to be chosen from restricted sets. One practical implication of this move is that the status quo may become a presumption of such modes of reasoning; as O'Neill comments, "the weighing of preferences will be limited by this initial listing of options, which itself precedes, and so derives no vindication from, instrumental reasoning" (BJ, 18). What this means is that result-based reasoning becomes reliant on preformed categories of thinking that themselves receive no justification from practical reason and so are in danger of being arbitrary categories.

O'Neill observes that a common response to this area of criticism is that although such reasoning does rely on subjective categories or preferences, the construction of social structures that allow "neutral" experts to evaluate preferences overcomes this potential limitation and eliminates the charge of arbitrariness (FH, 87). O'Neill suggests instead that even though such experts may have what she calls the "benevolent motivation" to act as a neutral observer, they may be unable to fully detach themselves from the established categories embedded in their cognitive capabilities (FH, 86). There are three general implications from this. The first is that the neutral categories they seek are still open to the charge of arbitrariness, as they remain unjustified by practical reason, and instead rely on pre-established agreement. Secondly, this observation implies that such an approach is over-ambitious about the capacity of rational agents like human beings to escape from the context of principles and social norms within which we are embedded. Lastly, this approach removes the possibility of criticising such established norms and categories of thinking, as they are adopted prior to the deployment of reasoning that remains instrumental (cf. FH, 92).

O'Neill notes that while this may not turn out to be a practical problem for "commercial or public policy reasoning", it becomes problematic when we attempt
to deploy such reasoning in relation to ethical problems. This is because, “at most such reasoning could reach conditional conclusions”; any proposed guidelines for action would be contingent upon the acceptance of certain principles or norms that are left unvindicated (BJ, 18). The most significant implication of this for O’Neill is that result-based reasoning seems in the end to become reliant upon “action-based and frequently norm-directed patterns of practical reasoning used in daily life and in institutional settings” (BJ, 18). The paradoxical conclusion seems to be that instrumental reasoning about ends leads inevitably to a form of reasoning that begins with action. The objectionable implications of this are, firstly, that institutionalised patterns of thinking remain unquestioned, with no resources available for their critique, and, secondly, that justice can be understood in no other way than in subjective terms. This reduces justice to an understanding in Utilitarian terms; as O’Neill phrases it, “utilitarian deliberation makes justice nothing other than basic beneficence” (FH, 95).

Leaving aside these serious considerations regarding the difficulty of establishing the “neutral” terms required for instrumental reasoning about results, we may now turn to the more fundamental problems O’Neill finds in the reliance of instrumental and result-based reasoning on an account of agency that assumes as its foundation a causal connection between preference and agency.

### 3.1.2.2. O’Neill’s critique of linking action to “preference”

O’Neill distinguishes between realist approaches to preference, which regard them as “real states of agents at particular times”, and the “revealed” preference approach. This latter approach sees agents’ preferences as revealed through action, and assumes “that their preferences are systematically structured, for example, that they are connected, transitive and commensurable” (BJ, 16-7). This account of agency is deeply problematic for O’Neill. She argues that such an approach binds preference and agency together and as a result creates a model of agency that is ill-equipped to offer any substantial account of autonomous action,
i.e. one that is capable of demonstrating how moral reasoning relates to autonomy, or why autonomy should be viewed as worthwhile.

Firstly she claims that one immediate implication of this approach is that "agents cannot, by definition, act counter-preferentially... [and b]y the same token, cannot act counter-rationally" (BJ, 17). Agents instead are bound to their preference; all action reveals coherently structured preferences or desires. She observes that "[a]ction can then be criticized as irrational only when based on mistaken belief or calculation" (BJ, 17). Reason under this account is transformed into a limited and reduced capacity for determining courses of action with little or no influence from motivations other than the desire to produce an outcome that, for her, remains arbitrary or lacking justification. Secondly, she argues that an account relying so heavily on preference will find it extremely difficult to demonstrate the value of particular action based on particular preferences, over against alternative actions and outcomes. To highlight the significance of this objection she describes the varying attempts to link such accounts to a substantial account of autonomy, and concludes that none is readily available through a preference-based approach to action, as the subsequent section will show.

3.1.2.3. The failure of a preference-based account of agency to reach the level of principled autonomy

O'Neil comments that "the problem for an account of autonomy that starts with an empiricist model of action is to show why some but not all instrumentally rational ways of pursuing given preferences in the light of given beliefs should count as autonomous" (BJ, 33-4). Such attempts fail to justify action, since they fail to reach the level of normative autonomy. She describes three variations of attempts to identify autonomous action related to a preference structure. Firstly, there is autonomous action as "independent" action (BJ, 29). She asks, "can just any independent or unconventional action, including deviant and criminal activity, count as autonomous?" (BJ, 31) Autonomous action, understood as simply action independent from some particular thing, remains intrinsically arbitrary; no value
can be determined from such action. Secondly, she discusses the attempt to define autonomous action as some form of “rationality” or “coherence” in action (BJ, 29). While independent action cannot claim value of any kind, rationality or coherence may have a stronger claim to identifying value in action, given that they prize what is at least seemingly reasoned action. However, she comments that “interpretations that stress rationality or coherence may suggest, at least vaguely, that... [autonomy] is to be valued, but do not show what makes it distinctive, or whether it is in any way linked to agency, freedom and independence” (BJ, 31). We must again ask what is it that makes reasoned action valuable, surely even the most deplorable acts, or the most dependent action, can reveal reasoned thought processes and coherent structuring? Again we see that such a justification for autonomous action remains neutral in respect to any attempt to distinguish permissible from impermissible forms of action.

O’Neill also discusses the ways in which empiricist accounts may combine independence, rationality and coherence in attempts at identifying the value, or autonomy, of certain action. She rejects John Stuart Mill’s claims that the content of action could be graded in relation to higher or lower pleasures; using the examples of “poetry” and “pushpin” she shows that such a method fails to indicate the value of one over the other, again revealing the action-guiding capacity of preference as bound to seemingly arbitrary whim (BJ, 34). She rejects also the ordering of preferences, again on the basis that such coherence in action is in fact dependent on some account of a “substantial self”, or of “free will”, as it relies on the idea of “endorsement”, which we must interpret as implying some deeper capacity for autonomous agency if it is not be understood “entirely formally” (BJ, 35). She also rejects this justification on the basis that we could identify coherent links between almost all immediate or lesser preferences and preferences of more substantial content, and such links often reveal no form of autonomous agency; e.g., the decision of a child to rebel against parental demands creates a link between this substantial preference and more trivial acts, yet nothing of autonomous quality can be identified from such a situation (cf. BJ, 36). O’Neill concludes from these observations that it may in fact be the case that independent action and rational or coherent action are intrinsically incompatible; independent action is entirely
neutral on the subject of value, and reasoned action, while potentially evidencing some form of value, has no basis for claiming to be autonomous action (cf. BJ, 39-40). Ultimately then, she rejects the notion that either demonstrate any distinctive form of autonomy.

Thirdly, she rejects a conception of autonomy that sees it as an agent’s capacity for choice, or what she terms “mere, sheer choice” (BJ, 38). She uses an evocative piece from Iris Murdoch to suggest that such a being could potentially be in no way distinguishable from the sinister character of Lucifer (cf. BJ, 38). The central objection she makes to this argument is that while at its root it is dependent upon an account of the free agency in human beings, it is also relies on an empiricist account of action that is limited to the causal link between preference and action; therefore “there is no room... for the more radical existentialist view of human autonomy” (BJ, 39). For such an account to explain the value of action it would need a deeper account of autonomy, one it is conceptually obligated to reject.

O’Neill therefore concludes that preference-based approaches to action fail to identify any action as meeting moral standards, and this she finds to be a fatal defect for such accounts. Two practical examples employed by O’Neill demonstrate what is at stake. In the case of an issue as urgent as global hunger, where one would assume widespread agreement among result-based reasoners, instead we find that such reasoning deploys other potentially arbitrary categories to provide the justification for their humanitarian efforts; as she puts it: they rely upon “whatever other (ethical?) outlook provides the assumption that the destitute care most about basic needs” (FH, 67). Given that such reasoning accepts only an empirical causal account of agency that links it to subjective preference, it must look elsewhere for the vindication of moral imperatives. Secondly, in terms of global economic justice, she observes that consequentialist reasoning presupposes “that the underlying context in which policies and decisions that affect ‘aid’ and trade are made is the existing international economic order” (FH, 62). So not only must instrumental reasoners remain silent on the criteria for morality, but they must also remain silent in the face of systems and structures which stand in need of revision and criticism.
In summary, it is O'Neill's contention that proponents of instrumental reasoning, in their attempt to avoid an account of human freedom or agency, in fact end up with few if any resources to answer questions about how and why autonomy is relevant for morality or justice. On both a theoretical and a practical level O'Neill argues convincingly that instrumental reasoning and empirical accounts of human action, that reject accounts of human freedom and agency which rely on a more substantial account of autonomy, are incapable in the end of offering any meaningful guidelines for just or moral action. This allows O'Neill to offer an alternative, potentially more substantial approach to thinking about the role of moral agency or autonomy in matters of justice. An understanding of her alternative Kantian approach must begin with her interpretation of Kant's conception of human freedom, as it is the foundation of his account of agency, reason and moral autonomy.

3.1.3. A Kantian account of human freedom

O'Neill's Kantian project is dependent upon the account of human freedom he sets forth. This analysis allows us to adopt a particular standpoint which Kant, and O'Neill, consider crucial to understanding morality in the context of our capacity for reason, and for understanding the relationship between free will, autonomy and morality. In this section I will briefly summarise how O'Neill suggests we should interpret Kant's approach to freedom. I discuss her emphasis on the importance of relinquishing a separation of, on the one hand, the idea of freedom as a noumenon, from, on the other hand, observable phenomena, as if they existed in ontologically separate worlds. The intellectual inconsistency of proposing theoretical knowledge while rejecting an account of human freedom is also noted.

As to the theoretical status of "freedom", an account of the Kantian concept of freedom as understood by O'Neill must begin with one crucial consideration. As O'Neill has emphasised, Kant's account is to be recognised as a defence, or a vindication, and not as a proof, of human freedom (cf. BJ, 42). Like for the idea of
God, there can be no proof of an idea that cannot be demonstrated as an existing entity that the senses can confirm. As we will discover, this is significant because it is the very question of the scope and limitations of theoretical evidence that is at stake; the entire vindication of freedom rests on the assertion that an empirical perspective of the world is incomplete. What Kant proposed instead was that there was a complementary approach in looking at the world, specifically at ourselves, and that this approach encompassed that which could only be denied by a wholly empirical account, in particular human freedom. He argued that we could, and must, adopt an alternative, practical perspective, one which recognised the capacity of our will to freely determining moral action. In this way his account is not a proof of freedom but an argument for why we should understand ourselves as free for practical purposes (cf. CR, 55). Indeed O'Neill notes that it is “in acting we must take the practical standpoint” (BJ, 48).

O’Neill identifies two aspects to Kant’s understanding of human freedom, namely positive and negative freedom. Negative freedom can be understood simply as a freedom from extrinsic authority. She writes, “[h]is account sees negative freedom or freedom of the will as a capacity ‘to work independently of determination by alien causes’ and positive freedom or autonomy as a specific, coherent and reasoned way of using negative freedom” (BJ, 43). O’Neill is attempting to defend Kant against the objection that his claims are “metaphysical”, in this case, that his distinction between the noumenal and the phenomenological realms posited the existence of ontologically separate worlds. She argues instead that Kantian negative freedom should not be interpreted as implying a self that is wholly detached from nature, and outside the causal order. Indeed, Kant insisted that action could and should be viewed as caused, in the theoretical sense. Instead, negative freedom must be interpreted as meaning that our will is not governed by empirical causation. We are indeed within and a part of nature yet we possess cognitive and volitive capacities that allow us to see ourselves as other than and apart from these natural sequences. It is important to note that here, in this explication of negative freedom, Kant introduces what he termed “heteronomous” authority, which he sees as opposed to autonomous authority. Heteronomy is the sourcing of authority from any external source, and he understood “external”
broadly, as implying anything from one's own desires and inclinations to the authority of political groups or religious organisations. For Kant one could never be understood as free if one submitted oneself to any form of heteronomous authority.

It may seem like a contradiction to suggest, first, that action can be understood theoretically, and then suggest that our action is not ruled by natural causality. To clarify this we must view the noumenal and the phenomenal as two perspectives from which we may look at action. These two perspectives are each indispensable for the agent. The theoretical perspective allows us knowledge and experience of the world we inhabit (although this experience is only of phenomena, of the "appearance" of things, not things as they really are) (cf. CR, 59). The practical standpoint does not allow us theoretical knowledge of things, but it does allow us to deliberate and reason about action (BJ, 46). It allows us to propose act-descriptions and apply reasoned judgment to them. Deliberation of this kind epitomises the cognitive capacity that Kant suggests is the essence of the practical standpoint. This perspective is doubly indispensable, indispensable if we wish to reason about action in the world, and also indispensable if we wish to undertake empirical investigation of the world we inhabit. We cannot hope to arrive at empirical, evidence-based conclusions on the nature of things without recognising that we have the cognitive ability to undertake such a task. O'Neill remarks that Kant's argument amounts to the assertion that it is impossible "to argue freedom away": we cannot question the nature of things unless we are free to do so (CR, 63).

Again we note that this is not a proof of the existence of human freedom. O'Neill describes in her Tanner lecture on religion and reason how Kant proposes that we "postulate" this freedom. He claimed that practical reason requires that we "postulate" a free self. On this understanding freedom is a "hope"; taking the starting point as a practical, political attempt to answer questions relating to knowledge, action and hope, we can legitimately see ourselves as free in the sense

---

162 Onora O'Neill, "Kant on Reason and Religion", The Tanner Lecture on Human Values (delivered at Harvard University, April 1-3, 1996)
163 O'Neill, "Kant on Reason and Religion", p. 286.
that we distinguish our capacity to reason and act from the causally ordered world which we encounter through our senses. Also the very cognitive actions involved in theoretical experimentation and phenomenological knowledge gathering require us to “postulate” our own freedom.164

In summary, O’Neill interprets Kant as arguing that while we can never demonstrate freedom in any empirical or theoretical sense, we can and must assume a level of freedom attributable to human agents, as this is the only way we can coherently understand the world, and deliberate about our action in the world. We must simultaneously see natural causality as a dominant feature of the world we inhabit, and yet see ourselves as capable of determining our own action. To do this all that is required is that we shift from the empirical to the noumenal perspective. While Kant has ruled out any theoretical proof of the existence of freedom, he gives us reason enough to assume, solely for the practical task of collaborating to construct principles of universal validity, that human beings, as finite and rational beings, have a capacity for free agency.

However, as O’Neill points out, the possession of negative freedom, according to Kant, is not of itself sufficient to meet the requirements of free agency. It is indeed freedom from things, whether they be our own desires or other persons’ wills, but it is not yet freedom in action. For Kant it is our capacity for positive freedom that is so significant, for it is this that allows us to be moral. On a purely negative view of freedom the connection between reason and action is not evident. O’Neill explains that “freedom cannot be merely negative nondetermination by ‘alien’ causes, but must include the capacity for self-determination or autonomy; the exercise of autonomy, in thinking and in acting, is what we call reason” (CR, 69). We must conceive of a positive form of freedom, where we apply practical reason to action, in order to make our action intelligible. In this way we may understand ourselves as free, i.e. as capable of intelligible, specifically moral, action (cf. CR, 60). The crucial conceptions behind the Kantian account of positive freedom are practical reason, autonomy and agency. In what follows I will explore O’Neill’s reconstruction of the formal expression Kant gave to these conceptions, the

different versions of the “Categorical Imperative” he establishes as the measure of morality.

3.1.4. O'Neill’s interpretation of the formulations of the Categorical Imperative as interdependent realisations of human freedom

As we saw in the last section, Kantian freedom has both positive and negative dimensions, and it is in the positive use of freedom that we begin to understand the foundations of Kantian morality. What is most distinctive in the understanding of Kantian positive freedom, and therefore Kantian morality, is that Kant saw free action as autonomous action. This idea links together the Kantian conception of human freedom with his interdependent accounts of practical reasoning and autonomy (understood as positive freedom or moral agency), and allows us to attempt a reconstruction of what the positive use of freedom entails, namely the adoption of the Categorical Imperative as the method for evaluating act-descriptions and determining moral action. In what follows I will highlight those formulations of the Categorical Imperative that are most relevant to the work of O'Neill, and also attempt to show how autonomy, agency, freedom and reason can be accounted for within this structure. I will conclude with a summary of the connection between these four conceptions, which will reveal how they may be understood as related expressions of the basic human capacity for morality.

The Kantian approach adopted by O'Neill does not begin with any account of the good at which we may then direct the focus of procedures for arriving at ethical principles. Rather, she suggests Kant makes no assumptions about what we may want or desire, and only allows some basic assumptions about the nature of human life, such as that humans are “finite, mutually vulnerable beings” (B1, 79). On this basis Kant begins by asking what principles a plurality of agents, who aim to discover the rules which will allow them to live together with some measure of social coordination, must choose if they are to succeed in forming a mutually respectful and relatively peaceful society. In this sense we see that the question, “what ought I do”, relates to the effort of finite beings such as ourselves to
construct principles by which we may live together peacefully. This is significant, as it is this interest in the mutual realisation of freedom which supplies the criteria for the method of determining principles expressed in the Categorical Imperative. Given that the intended outcome is the coexistence of all agents, we may ask, “how must our actions be structured or restricted to ensure that outcome?” As we have seen in O’Neill’s emphasis on Kant’s interest in the realisation of human freedom and the conditions required for this interactive, or as O’Neill says pointedly, “political” task, this is the essence of the Kantian principle of universalization; it is a response to this fundamental question.

The answer to the question, “what ought I do?” (or, put another way, “how may I discover what I ought to do?”), is presented by Kant in several forms, all of which are deeply interconnected, and this should be understood as reflective of the interrelationship of his conceptions of reason, freedom, agency and autonomy. Respect for persons (and therefore their agency) and the principles of reason and autonomy are the foundational ideas behind his formulations, the Formula of the End in Itself, the Formula of the Universal Law and the Formula of Autonomy. As O’Neill has highlighted, these are to be understood as different formulations of the same guiding concepts, though she notes that the Formula of the End in Itself is more fundamental and more demanding, as it involves the protection of human agency and the respect of persons (cf. RPR, 24). They can also be seen as different perspectives upon the same guiding principle. An exploration of the structure of these formulations will yield an understanding of their shared origins.

In what follows, I will attempt to summarise the structure and content of the aforementioned formulations. This will elucidate both the specificity of each, and the relationship between the three; how the Formula of the End in Itself, in its protection of agency, is simply an alternative perspective reflecting the demanding restrictions placed on reason by the Formula of the Universal Law, and how the Formula of Autonomy mirrors both in its defense of the capacity of reason to

---

165 O’Neill, “Rationality as Practical Reason”, p. 23. Further references in text (RPR). There are also several other formulations, which I will not explore here, as they do not bear on an understanding of O’Neill’s theoretical framework to the extent that the Formula of the End in Itself and Formula of the Universal Law do.
provide its own critique, and therefore to govern itself. In this way we shall be able to understand and describe the interdependence of Kant’s various expressions of realising human freedom.

3.1.4.1. Formula of humans as ends in themselves: agency and reason

The Formula of the End in Itself is a requirement that we treat others as ends in themselves. What this implies is that we refrain from instrumentalising others without their consent. The use of others as means is often unavoidable in everyday life, however, when we do so we must respect a double requirement, firstly, that the other in some way acknowledges and agrees to be a participant in our intended actions, and secondly, and more crucially, that the other is capable of acknowledging and agreeing to be a participant in our intended actions. Here the central concern is capacity for agency. The fundamental requirement of the Formula of the End in Itself is that the other’s capacity for agency is not intentionally impeded and that the option for consent is left open as fully as is possible in the particular context or situation. In her encyclopedia article, “Kantian Ethics”, O’Neill highlights what is so significant about the Formula of the End in Itself’s requirements relating to agency. In terms of the relationship between our capacity to act and the appearance of consent, she suggests that Kant’s approach is vastly different from most other definitions of this relationship. She claims that some others have suggested that we first posit a capacity for cognitive rationality that may or may not be true of all human beings, and on that basis prescribe the definition of consent, while others still have suggested that we can base a definition of consent on any consent sought and received from real persons (she calls this the “historically contingent consent of actual others”)167. Kant on the other hand suggests that what is of fundamental importance in matters of consent is the possibility of consent, or the individual’s capacity for consent. Whether or not an ideally rational being would consent to terms, or whether or not any person

---

166 Agreement taken as a reflection of autonomous agency, autonomous in the sense previously elaborated, and agency in the sense discussed previously, which will be discussed in what follows. Agreement is not to be understood as consent under any form of coercion or manipulation.

has actually consented to terms, is irrelevant at this level. What is relevant is whether or not a human being has in principle the capacity to consent, and that all the actual conditions necessary for the possibility of consent are met.\textsuperscript{168}

What is required then if we are to refrain from instrumentalising others without the possibility of their consent? We have seen that it requires that consent be possible, and we may reformulate this as the requirement that the other's capacity to act, i.e., to consent or dissent, is left intact.\textsuperscript{169} Agency is the basis of consent, and so we must adopt principles by which agency can be respected. This means that we must reject principles that make it impossible for others to carry out the actions that we ourselves may wish to undertake. We must not will an action and at the same time will that this act be made impossible for others (cf. RPR, 24-5). O'Neill notes that "[w]e treat others as ends by acting in ways in which a world of agents can be sustained: by acting on maxims that can form part of a system of maxims that can be willed without contradiction, that harmonise with the necessary conditions for sustaining a world of agents" (RPR, 25). The Formula of the End in Itself therefore is one way in which our efforts to seek principles by which we may live together can be expressed and fulfilled.

In this way we begin to see the connection between the Formula of the End in Itself and the Formula of the Universal Law. This latter formulation requires, through procedures of reason, that we adopt only maxims that can be universalized, i.e. that can be adopted by all persons. O'Neill specifies that "[t]he fundamental difference between the Formula of the Universal Law and the Formula of the End in Itself is that the former constrains what agents should (may, may not) do, whereas the latter constrains how agents should (may, may not) be treated" (RPR, 23). The Formula of the End in Itself addresses the significance of respecting the agency of others, while the Universal Law addresses the question of what we should do with our own agency.

\textsuperscript{168} As we have seen this has important implications for accounts of justice such as Rawls's that ground justification in consent, whether actual or hypothetical. It also has consequences in the area of medical consent, to which O'Neill has contributed extensively. See O. O'Neill and Neil C. Manson, \textit{Rethinking Informed Consent in Bioethics} (Cambridge: Cambridge University Press: 2007), and O. O'Neill, \textit{Autonomy and Trust in Bioethics} (Cambridge: Cambridge University Press: 2002), pp. 28-49.

\textsuperscript{169} O'Neill, "Kantian Ethics", p. 178.
3.1.4.2. Formula of Universal Law (and the supreme principle of practical reason): the conditions for the discovery of universal principles

Whereas the Formula of the End in Itself asks what principles must be rejected if we are to avoid instrumentalising others, Kant arrives at the Formula of the Universal Law by asking the analogous question: what principles must be rejected by a “plurality of noncoordinated (potential) actors and thinkers”, or, the enactment of which principles might damage social coordination, or might not contribute to a better coordinated society? (cf. CR, 16) Kant seeks to vindicate these restrictions by offering a justification, rather than a proof, and he seeks to justify his principle unconditionally, or to an unrestricted audience (cf. CR, 16).

Any justification directed at a restricted audience, Kant regards as “heteronomous”, because it is reason grounded in something external to itself. For Kant this has two significant implications. Firstly, that such reason will not be in principle universally accessible, because it is dependent upon the assumptions or starting points of particular groups, and therefore on contingencies that may or may not be relevant to all others, and, secondly, it implies that reason itself cannot claim validity or legitimacy. We will return to the latter implication when we look at Kant’s Formula of Autonomy, for now we will remain with the former, that of universal accessibility at the level of principle.

So what then counts as a valid maxim for a plurality of “interacting agents”? The only requirement, according to Kant, is that the maxims one proposes must be at the very least intelligible to one’s audience. Given that he is seeking validity directed at the entirety of a plurality of interacting agents (in the case of human beings, this implies the whole world of human beings), we must reject all maxims that can be valid for some and not others. So the Formula of the Universal Law then becomes the principle of seeking and adopting only maxims that can (in the modal sense) be adopted by all others, i.e. maxims of universal accessibility in principle.
Practical Reasoning and Transnational Justice

While this may seem like a purely formal requirement with no practical application, its structure is highly restrictive. The Formula of the Universal Law involves a double requirement, that only by willing maxims that leave the possibility for their enactment by all others intact, and at the same time leave the possibility of their continued enactment intact, may we arrive at maxims that meet the conditions of the task of discovering ethical principles. The latter requirement, that these maxims allow for their own continued enactment, is significant in that it clarifies a potential misunderstanding of the Formula of the Universal Law. It is not simply a requirement that we choose acts or act-descriptions that all others also choose (cf. RPR, 11). It also requires that the means to enact such maxims are left intact. For Kant it is incoherent to will a particular end while at the same time willing the impossibility of that end (cf. RPR, 12). This has two significant implications. Firstly, it places greater restrictions on the maxims that would pass the test of universalizability. No maxim that contains within itself its own destruction can be permitted; a standard example is that of a principle of deception; deception simultaneously destroys the means for deception to be possible, i.e. it destroys trust (cf. RPR, 12).

Secondly, O'Neill has observed that the Formula of the Universal Law "supersedes" the theory of instrumental rationality: its structure includes that we also will the means to our ends (BJ, 77). Means-end logic is an intrinsic and essential component of the Formula of the Universal Law, though it is a subordinate aspect within the requirement of universalizability. As we have already noted, the limitation of instrumental rationality is that it is inherently arbitrary, whereas the Universal Law rejects all forms of heteronomy in ethics (cf. RPR, 9). This means that it rejects any authority for reason that is external to reason itself, and demands that reason be autonomous. In this way we see how instrumental reasoning is incomplete or inadequate when compared to the Universal Law, which incorporates both means-end reasoning and a capacity for moral decision-making.170

170 To fail to act in accordance with the Formula of the Universal Law is to treat oneself as an exception, for you intend both that the means to your ends be preserved, and that you may simultaneously undertake action which would disturb those ends were it carried out by everyone. (RPR, 12)
The Formula of the Universal Law can be understood both as a formula for guiding action and as a formula for regulating reason that wishes to appeal to “the world at large”. In this way it is equivalent to the “supreme principle of practical reason”: those restraints we place on our maxims we also place on our reasoning (cf. RPR, 16). We should not and cannot offer modes of thinking to an unrestricted audience that cannot be adopted by that audience. If we wish to reason to the whole world, we must at least hope and assume that our reasoning will be “followable” and “intelligible” by all (RPR, 16. Emphasis mine). Only reasons that contain principles and acts that could potentially be acted upon by all others are truly followable and intelligible to others. Anything less would imply an appeal to heteronomous authority, and such a method for the legitimation of reason is unacceptable to Kant, as it would not be universally accessible. Instead, Kant argues for the autonomy of reason, and we shall now investigate what is required for reason to be autonomous.

3.1.4.3. Formula of Autonomy: a recursive and reflexive critique of reason

We have returned to the question of heteronomy in ethics, and therefore it is now appropriate to discuss how and why Kant locates the authority of reason within reason itself. As previously noted, O’Neill emphasises that the relationship between reason and autonomy is fundamental to Kant’s ethics. When Kant speaks of autonomy, he refers to the method by which we regulate our capacity to reason. By reason, and by critique of reason, he refers to the recurrent subjection of reason to its own laws, to no other authority besides its own (cf. RPR, 27-8).

O’Neill argues that Kantian autonomy should not be understood in any of the traditional ways in which it has been understood. As we have seen, in Bounds of Justice she gives a detailed account of the interpretations of autonomy she wishes to reject, specifically autonomy as individual independence, rationality or coherence, or “mere sheer choice”. What she identifies as being significantly different in Kant’s account of autonomy is that “Kant never writes of autonomous
Practical Reasoning and Transnational Justice

selves or persons or individuals" (RPR, 27). Rather, he "predicates autonomy of reason, of ethics, of principles, of willing" (RPR, 27). This concept of autonomy is to be understood by its literal and traditional definition, that of self-rule, and crucially as the self-rule of reason, not of individuals as such. It does not imply that we are free to arbitrarily choose personalised laws by which we shall live; rather, we submit our reason to its own evaluation and accept the law-giving capacity of our reason.

For Kant, reason must govern and evaluate itself. To discover how reason may do this (and also how we may avoid claims that reason governing itself is an arbitrary requirement which can lead us nowhere), we must ask what minimal requirement a reason would need for legitimacy or justification. To answer this we must return to Kant's initial purpose, that of discovering principles for a plurality of interacting agents, about which he makes no assumptions other than those previously mentioned, such as finitude and interdependence. We must also take into account that this plurality is to have no pre-established agreements. This leads us back to the insight we have already explored in relation to the Formula of the Universal Law and the principle of practical reason. For reason to be judged legitimate, it must be accessible and intelligible to all, and must provide principles and arguments that are adoptable by all.

The fundamental concern, again, is agency. To follow, understand and adopt proposed reasons for thought or action we must be at least capable of following, understanding and adopting principles or reasons. In other words, we must have some capacities for agency that allow us undertake action, such as adopting or accepting a principle or reason. Legitimate reasons are those that do not seek to suppress or undermine agency, but rather respect and support the capacity to understand, follow and possibly accept or endorse reasoning. On this understanding, reasoning that seeks to undermine agency is not legitimate, as it is not accessible to its audience. Examples of act-types that undermine agency, and so neither respect agency nor offer legitimate reasoning, are coercive, deceptive or violent act-types. This idea can be expressed as universalization where the audience is understood to be unrestricted. And as we have seen, Kant argued that
only unrestricted reasoning could be autonomous, i.e. free from heteronomous sources of authority (such as the shared agreement of a particular audience). We are therefore not autonomous when we act in ways that express varying levels of independence or rationality, but when we apply “reflexive”, recurrent evaluation to our maxims, or our practical reasoning about action (RPR, 28).

For Kant, the self of self-legislation is that of reason, of reflexive application of the principle of universalizability to one’s reason. We are autonomous in no other sense than that in which we have the capacity to determine and act upon maxims which adhere to the self-legislating demands of reason. The root of our autonomy is our capacity to derive law-like principles from practical reason. Reasoning autonomously therefore, for Kant, is the most fundamental way in which we express our freedom.

In summary, these three formulations express the interrelationship between Kantian agency, autonomy, freedom and reason. What is evident from an analysis of this connection is that there is an almost symbiotic dependence between these conceptions: none can be analysed or interpreted in isolation. The Formula of the End in Itself expresses the demand for respect of persons, and therefore of agency. This is accomplished by the restriction of action to acts reflecting maxims, which are, and will remain, universally performable. This is an alternative expression of the method by which the Formula of the Universal Law restricts maxims to those which can be universally adopted. The Universal Law, as the supreme principle of practical reason, is grounded in practical reason. It asks how must reason be regulated for the accomplishment of a practical, social project. We begin to see how agency and reason interact, however we cannot fully grasp either without including the notion of autonomy. The Formula of Autonomy states that reason is only autonomous if it rejects all external influence and submits itself to reflexive evaluation, i.e. to its own evaluation. Autonomy is therefore the self-governing capacity of reason. Free agency is simultaneously protected by, and expressed through, the autonomy of reason. Positive freedom in the Kantian account can therefore be understood as action performed freely, i.e., autonomously, through a commitment to the self-legislating potential of reason. Autonomous free action is
reasoned action that respects the agency of all others, refrains from instrumentalisation and leaves open the possibility of comparable free action for all others.

3.1.5. Conclusion

What has hopefully become apparent from my discussion of O'Neill’s philosophical framework is the advantage of her consistent approach within Kantianism. Rather than avoiding the difficult issues of human freedom and autonomy by adopting an approach to rationality and action that allows only a reduced conception of agency and autonomy and denies a conception of human freedom, O'Neill ambitiously adopts a coherent Kantian system for integrating agency, autonomy, reason and freedom. This account is both robust and practical; it offers a sound theoretical basis for rejecting preference-based approaches to agency as well as instrumental, result-based reasoning, and justifies itself by appeal to practical considerations regarding the coordination of shared principles for moral agents. Perhaps the charge still remains that it does not provide an adequate justification on empirical parameters, though one would imagine that Kant, or O'Neill, would respond that theoretical justification is unavailable as such empirical arguments are not equipped to provide vindication for conceptions which are necessarily beyond the realm of empirical evidence and facts, and which offers a critical perspective that enquires into their constitutive elements, conditions and grounds of validity.

All the formulations of the Categorical Imperative, and therefore also the conceptions of agency, autonomy, freedom and reason, derive from the initial premise: the practical, political purpose of constructing and discovering "modes of coordination" for a plurality of agents (CR, 18). Human autonomy is an expression of our capacity to adopt principles that contribute to the establishment of just ways of life for human agents; this is how Kant demonstrates and vindicates our rationality and our freedom. This approach successfully argues for a substantive account of freedom that includes only a limited and minimal account of human attributes, something we noted was necessary when we looked at O'Neill’s
criticism of the unnecessary idealisations and circularity inherent in Rawls's theories of domestic and international justice.

Given that this approach centers on the discovery of principles for action, it is therefore to be understood as a duty-based approach that emphasises obligation as its starting point. As we have seen, O'Neill argues that such an approach is more valuable and productive for questions of justice than an approach that only emphasises the rights of the individual. In terms of the critique of institutions, O'Neill's Kantian approach not only offers greater resources for revision, critique and reconstruction, it also highlights the important role that institutions and collectivities can play in international justice. It also offers strong reasons for rethinking distributive justice. O'Neill suggests that one of the implications of the Categorical Imperative is that any principles that rely on coercion must be rejected. This can only be accomplished by avoiding differentials of power that induce persons to accept "offers they can't refuse".

With O'Neill's framework, we are not governed by subjective desires that fail to provide resources for thinking about principles of justice that are coherent and compatible for a social union potentially as large as the whole world. O'Neill observes that "[r]eason and justice are two aspects to the solutions of the problems that arise when an uncoordinated plurality of agents is to share a possible world" (CR, 16). We may understand reason and justice as being interlinked, with practical reason providing the framework for developing principles of justice.

The preceding considerations allow us to now review O'Neill's critique of Rawls's account of public reason and her defense of an alternative Kantian approach to public reason that emphasises universalizability over reciprocity. I will argue that the principle of toleration, so central to Rawls's account of international justice (and his limiting of human rights), receives greater justification from a Kantian approach to public reason that takes the conditions for the possibility of

---


172 See O'Neill, Bounds of Justice, Ch. 5. "Which are the offers you can't refuse", pp. 81-96.
communication, seen as a form of agency, as its starting point. Two broad conclusions from our critique of Rawls's account of justice are relevant to our analysis of public reason. Firstly, that the Rawlsian approach can only be justified by appeal to liberal democratic values that remain unvindicated. Secondly, that the Kantian alternative offers strong reasons for rejecting all forms of reasoning about justice that draw unjustified distinctions between persons within or outside of certain boundaries.

3.2. O'Neill’s interpretation of Kantian public reason

Having discussed O’Neill’s Kantian foundations, in particular the equivalence between the separate formulations of the Categorical Imperative and their shared derivation from the conditions for the possibility of reason and action amongst a plurality of finite and vulnerable agents, I now address O’Neill’s distinctive interpretation of Kantian public reason. I will begin by reviewing Rawls’s approach to public reason (3.2.1.). Rawls proposed that public reason be grounded in an overlapping consensus on a set of political values that “could”, by their neutrality toward ethical worldviews, be endorsed by all “reasonable” comprehensive doctrines. The corollary of this account, as was discussed in Part 2, is that justice and justification are then restricted to the domestic society.

O’Neill argues that Rawls's account of public reason is in fact, (in comparison with Kant’s understanding, a form of private reasoning (3.2.2.). Any form of reasoning that appeals to pre-established agreement, rather than vindicated premises, to ground itself, appeals to an external and therefore heteronomous source of authority. Shared premises may be constructive for those who share them, but are inaccessible to those who do not begin from such a consensus. O’Neill argues that shared political identities are not assumptions we can make about modern liberal democratic societies, and that the boundaries of the contemporary world are too porous for domestic societies to affirm private forms of reasoning. She also suggests that relativism is neither promising nor convincing as an attitude toward
reasoning across traditions and cultures, and that we must reject relativism if we wish to avoid a choice between silence and coercion.

Following from a rejection of relativism, I discuss David Hollenbach's critique of restricted or relativised forms of public reason, and his alternative account of "intellectual solidarity" between different traditions, including religious ones (3.2.3.). Hollenbach's account, based on a theological ethical approach that takes the "common good" as oriented towards human rights, offers the important insight that rights presuppose and are hence not threatened by deliberative processes; that the common good and human rights are mutually supportive. He shares with O'Neill the insights that abstract principles can only be fully realised when applied to particular contexts, and that free discourse is not a threat to basic rights or principles of justice, as these must be presupposed by accounts of public reason, deliberation and communication.

O'Neill's Kantian account argues that, in order to prevent the imposition of heteronomous sources of authority, public reason must be free, open and critical. For this reason O'Neill asserts that toleration is not an ideal but a necessary condition of public reasoning (3.2.4.). Rawls's understanding of tolerance as avoidance does not facilitate the dissent and criticism necessary to ground the authority of reason, and it therefore cannot claim objectivity. The proper object of toleration is communication; toleration applied to communication is demanding and involves listening and responding to others.

Although freedom in reasoning is essential, radically free modes of discourse cannot sustain themselves, and therefore reasoning requires some constraints (3.2.5.). The guiding consideration behind the constraints on public reason from a Kantian perspective is that principles and reasons that cannot be shared by a plurality cannot guide reason and communication. Such a consideration is recursive and modal, which implies that while reason constrains discourse, it is also established and vindicated through discourse. In this way, Kantian public reason is grounded in the Categorical Imperative and the supreme principle of practical reason.
O'Neil argues that a focus on rights, such as on freedom of expression, cannot adequately account for the constraints necessary to promote and enable communication. She argues instead that a focus on the obligations of communicators can offer a more comprehensive perspective by highlighting that some action can respect rights of expression and participation yet also undermine discourse and erode conditions for communication (3.2.6.). Accounts of toleration and communication must consider which sorts of action can undermine practices of communication. Certain constraints are necessary and are best understood as institutionalising practices that enable communication through the regulation of action that hinders it. O'Neill argues that respecting others' voices in free and open public discourse places great demands on communicators, who must "foster diversity" in communication. Tolerance ensures the shareability of reasons by testing them in discourse; practices that undermine this process are not tolerable. Which practices promote and sustain communication can only be discovered by applying these abstract considerations to particular communicative contexts.

Both toleration and the communicative obligations that stem from toleration are extremely demanding in the context of seeking agreement out of deep ethical plurality (3.2.7.). In this section I briefly review the political philosopher James Bohman's critique of Rawlsian public reason. Bohman argues that a free, open and engaged form of public reason is demanded in modern pluralist democratic societies. He highlights an apparent tension between principles of freedom and equality, and liberal values (based on fairness and reciprocity) in pluralist contexts, where liberalism is only one voice among many. Like O'Neill, he views discursive norms as "enabling conditions", rather than "constraints", which he associates with Rawlsian public reason. However, while Bohman derives the principle of toleration from democratic ideals of freedom and equality, O'Neill's account grounds principles of justice, public reason and tolerance in the same guiding conditions of the supreme principle of practical reason.

The constraints on discourse are moral constraints that apply to all practical reasoning about action (and therefore to communication): they offer a starting
point from which contextualised discursive practices can be established. In this way O'Neill argues that reason both constrains and is established in discourse. Her account promotes a public space for free and open deliberation between disparate perspectives, and restricts that space only where a lack of restriction would disable communication and therefore prevent agreement. In this way it reflects the Kantian foundation for the justification of principles of justice, and offers an accessible and shareable account for reasoning about justice, for deliberating about the common good and for debating questions of political identity, all in the context of porous boundaries and transnational interaction.

3.2.1. Rawlsian public reason intended as the shared reasoning of citizens

Rawls’s *A Theory of Justice* broke away from the philosophical trends of the most common approach to social ethics, Utilitarianism, and argued that a concept of justice that respected basic individual rights was the proper concern of social ethics. His account of justice rejected Utilitarian arguments in favour of principles that secured the distinctiveness of persons and prioritised the right over the good. He justified these principles by ultimately appealing to the considered convictions of moral persons who were members of a closed society.

As we saw in Part 1, Rawls had initially argued, (extending Kant) that there existed a congruence between the right and the good, i.e. that liberal principles would come to be accepted by the members of a liberal society as not only just, but also as part of their conception of the good. He suggested that a fundamental element of each person’s good related to the expression of his or her true nature as free and equal moral persons, and that his principles of justice enabled the “expression” of such a nature.

However, after *A Theory of Justice*, Rawls began to seriously doubt the efficacy of such arguments in eliciting support from the majority of the members of a liberal society. Instead, he came to believe that arguments grounded in assertions about the good of persons could never hope to achieve widespread support in a modern
Practical Reasoning and Transnational Justice

liberal society. In coming to this belief he shifted the focus of his work from the working out and justification of specific principles of justice to questions relating to what he termed “the fact of reasonable pluralism” (PL, 55). Rather than seeking to offer extensive justification for a specific set of principles, he now sought wider and specifically political justification for a family of liberal doctrines.

“The fact of reasonable pluralism” was for Rawls what constituted the modern context of justice. Modern societies could no longer be understood as ethically homogenous; the progression of history had brought with it a deep diversification of the range of moral, religious and philosophical worldviews that constituted modern societies. The legacy of the religious wars of the previous centuries had been the promotion of tolerance and a measure of acceptance toward the “other”, even if these only amounted to a modus vivendi.

All this implied for Rawls that questions of the good engendered deep division and conflict within society, and were best bracketed by a theory of justice that sought a path between conflicting doctrines. Rawls concluded that “irreconcilable conflict” was the inevitable outcome of interaction between diverse traditions and doctrines. He justified this conclusion by reference to what he called “the burdens of judgment”. This was his answer to the question, “Why does not our conscious attempt to reason with one another lead to reasonable agreement”? (PL, 55). Or, given that we share capacities for reason, why is it that the use of these powers in reasoned debate does not result in agreement?

The “burdens of judgment”, or the causes of reasonable disagreement, include problems relating to conflicting evidence, to disagreements over the weight of different considerations, and to differences in experience and the impact these have on our judgment and our interpretation of context (cf. PL, 56-6). These considerations imply for Rawls that there can be no guarantee of agreement even where discourse and debate has been reasoned and disagreement cannot be blamed on “rivalries for power, status and economic gain” (PL, 58). The free use of reason, for Rawls, had been shown to be ineffectual in the resolution of deep differences of perspective.

236
It was in this context, then, that he proposed an account of liberal justice that could be justified in purely political terms to an audience constituted of citizens of a liberal society holding diverse worldviews. By focusing on “the political”, Rawls sought to eliminate all traces of “the metaphysical” from his account; he abandoned what he interpreted as the Kantian elements of his earlier work in favour of strategies of justification that relied solely on the political context of a liberal democratic society.

Central to this new strategy of justification was Rawls’s conception of public reason. Based on his conclusion that reasonable disagreement could be expected from “free discussion”, Rawls argued that public debate in a democratic society could not hope to succeed unless it looked for another basis for “reasoned political agreement”. Citizens, as “reasonable” members of a liberal polity, must agree to abide by fair terms of cooperation and to offer reasons for action that all others could, or rather would, accept. For Rawls, this meant agreeing to only offer reasons in the public sphere that could be publically justified, i.e. to not offer arguments that depended on citizens’ comprehensive doctrines or worldviews, or at least to be willing to translate such arguments into those accessible to other citizens.

But by what procedure or by what conditions are reasons to count as public? While Rawls did not specify procedures for the public use of reason, it is clear that the content of this conception of public reason was to be found in the political context of a liberal democratic society. Specifically, it was made up of the reasonableness of the citizens of such a polity. Rawls came to the conclusion that agreement would only be “possible” (i.e. probable) if there was some shared basis for such an agreement, i.e. some pre-established coordination of values that allowed for an overlapping consensus on certain crucial political questions. Given

173 Rawls stressed that public reason related only to “constitutional essentials and matters of basic justice” (PL, 140); the wider public sphere was free to conduct debate on less restricted terms. 174 From Political Liberalism: “many of our most important judgments are made under conditions where it is not to be expected that conscientious persons with full powers of reason, even after free discussion, will all arrive at the same conclusion”. (PL, 58)
that moral, religious or philosophical values were too diverse and contentious to form the basis of such a consensus, the consensus must rely on solely political values. Therefore, as O’Neill has pointed out, it was the shared political identity of citizens, as fellow citizens of a liberal democratic society, which formed the content of Rawls’s account of public reason. Before they entered the arena of public political debate, citizens already saw themselves as sharing a common identity and certain core political values. It was this common understanding that made further political agreement both “possible” and likely. Pre-established consensus is the prerequisite of all further agreement in a liberal democratic society.

There are certain corollaries of this position. The most notable of which are that basic questions of justice can only be debated within societies that already share certain democratic political values, and that principles of liberal justice can only be justified to a restricted audience constituted by the citizens of such societies. This is not a radical reconstrual of Rawls’s position; his pessimism regarding the fruits of free and reasoned debate between different historical traditions implied for him that this limited justification of liberal values was the most that could be hoped for.

3.2.2. A critique of Rawlsian public reason as private and exclusivist reasoning

Kant drew a distinction between heteronomous and autonomous sources of the moral law. Heteronomy meant the derivation of authority from external sources, while autonomy was defined as the self-legislation of reason (in distinction from the individual self). On this account then, any forms of reasoning that rely on unvindicated premises are considered “private” reason. According to O’Neill, Kant

\footnote{175 "In assuming an idealized, closed society as the context of political justification, Rawls takes it that the public who are the proper audience for one another’s attempts at justification consists of fellow citizens, among whom there is in effect prior understanding that they form ‘a people.’ That is why he can speak explicitly of citizens as sharing a political identity (PL, xx), why he can designate the search for justice beyond borders as a search for a just law of peoples”. O’Neill, “Political Liberalism”, p. 419. And also: “[t]he conception of public reason advanced in Political Liberalism assumes not merely the shared sense of political identity of citizens of a closed society, but specifically the common sense of identity of fellow citizens in a democratic society”. O’Neill, “Political Liberalism”, p. 421.}
considered many public political offices to be private forms of reasoning, along with positions relating to churches and other associations whose reasoning relied on particular assumptions about the world. Any context in which the reasons offered would not be in principle accessible to those who did not already share certain beliefs or values was considered a private use of reason; such reasoning was internal to the thinking of a particular group and therefore private.

Evidently, under this interpretation, Rawlsian public reason – reasoning undertaken by a restricted plurality that already agrees on and shares certain political values – must be designated a form of private reasoning. The authority of such reason is contingent on unvindicated premises that are only shared by certain persons. What is crucial in describing a premise as “unvindicated” is not that others do not start from these premises, but that others are offered no reason for accepting such premises. In the case of Rawlsian public reason, the conceptions or “ideas” of practical reason which underpin the account receive their fundamental justification via the agreement or affirmation by certain persons. If the fact of agreement on certain principles, rather than the reasons for accepting certain principles, is fundamental to justification, then no reason has been offered to those who are not within this consensus to accept it as a premise. Whether or not there may be reasons (available on a different account) for accepting certain principles does not alter the fact that no reasons have been offered. If the most basic level of one’s account of reason relies on agreement, it is not possible for those outside of the consensus to accept its premises.

Such forms of reason that on closer consideration turn out to be exclusive are not only inaccessible to foreign outsiders; such premises may also not convince “citizens who stand back from the way things are, and ask whether they should be

---

177 This is not to suggest that for Kant there was no role for private reason, only that private uses of reason must not damage public reason: “However, Kant did not see any reasons to restrict (relatively) private uses of reason. On the contrary, he argues that the constitution of a just polity allows ‘the greatest possible human freedom in accordance with laws that ensure that the freedom of each can coexist with the freedom of all others.’ Even so, the difference between public and private uses of reason remains important... There are no good reasons for tolerating any private uses of reason which damage public uses of reason”. (PUR, 547)
that way”. Yet Rawls relies on the notion that the majority of citizens in a particular liberal polity will form a consensus. His account of reason for a liberal democracy can only succeed on the basis that citizens can be presumed to share a common political identity. However, O'Neill identifies the factors that make these assumptions problematic:

[It is all too well known that in real life persons are often unsure about their sense(s) of political identity, that they may find that those with whom they live in closed societies are not identical with those whom they regard as their own people, and that not all societies are closed. A central objective of politics may be the reconstrual of political identities, the separation or the merging of destinies rather than the working out of principles of justice to be shared within a closed society. Boundaries and identities are a central domain of thinking about justice rather than its fixed parameters. (PLPR, 420)

Rawls’s solution to the fact of reasonable pluralism appears to be inadequate to the real challenges presented by pluralism. It simply restricts the plurality of diverse worldviews to those who already share so much in common that further agreement is probable; shared political identity functions as a robust motivational factor in coming to agreement. If we could conceive of modern liberal societies as constituted by members who consider themselves a homogenous “people” with a shared political destiny, then this might not be so problematic (Although it would still leave questions of international or inter-societal justice unanswered). However, while “identity” (whether political, ethnic, religious or other) is a theme of discourses from different disciplines, both from individual sciences conducting empirical research and from philosophical approaches which oppose the reification of identity and point out the role of the “other” for the “self”, the main point is that identity must be a subject of justice, and not a presupposition:

[w]e cannot presuppose the contingencies of a particular social order or group in raising the most basic questions about the proper structure of society. We cannot presuppose the contingencies of a specific sense of political identity in asking the most basic questions about justice. (PLPR, 425)

---


179 The philosopher and ethicist Allen Buchanan also criticises Rawls, from a diagnosis of contemporary political realities, for proposing that shared political identity can ground an
Rawls's account of public reason is not a promising strategy in our contemporary context. As we saw in Part 2, such an approach endorses a relativist approach that rejects possibilities of communication between systems of thought from distinct cultures and traditions. This approach greatly reduces our options for constructive intellectual engagement in both the domestic and international public sphere. Of the inadequacy of a relativist approach to reasoning, O'Neill comments that

[i]f we do not bracket relativism, we have only two options for dealing with those whose ways of thought and life we do not understand. Either we can cut ourselves off and retreat to the cosiness of 'our' shared outlook; or we can impose our ways on others. If we are to have options other than quietism and imperialism... we must bet against relativism.\textsuperscript{100}

Such an apparent dichotomy is presented by Rawls's account of international justice: the choice between silence in the face of unjust political structures or else intervention and/or cultural imperialism. While Rawls is not communitarian, in that he does not think justice is defined by the distinct ethical traditions of particular societies, his account of international justice rests on the notion that justification is internal to particular societies, and that systems of thought that

understanding of modern societies. The focus of his critique is The Law of Peoples, however, it also applies to Rawls's account of public reason, which is the basis of many of the deficiencies in his account of international justice: "In Rawls's system, the fiction that the population of a state is a "people" unified by a single political culture removes problems of intrastate conflict from the domain of international law. The question of whether, and if so how, international legal institutions should respond to intrastate conflicts cannot even be raised within Rawls's framework because the parties who choose the Law of Peoples are understood to be representatives of "peoples," groups characterized by deep political unity and already possessing their own states. Rawls's political homogeneity assumption is not a mere detail; it shapes his understanding of what international law is for, by eliminating from the Law of Peoples principles designed to cope with important conflicts that arise within states." Allen Buchanan, "Rawls's Law of Peoples: Rules for a Vanished Westphalian World", Ethics, Vol. 110. No. 4. (July, 2000), pp. 697-721, p. 716. And further: "If the populations of states were, in fact, peoples in Rawls's sense, groups unified by deep agreement on conceptions of public order, then Rawls's talk of the rights of peoples to pursue their own conceptions of justice or the good without interference would make more sense. There would be no need to consider rights of groups within states. But the populations of states are not in fact peoples in Rawls's sense. They are, in almost every case, collections of groups, and in some cases the groups are distinguished by quite different conceptions of justice or the good. In addition, in many cases, groups within the state either deny that they should be within that state or they demand special group rights, including rights of limited self-government within the state, beyond the individual basic human rights that Rawls says international law should recognize. Buchanan, "Rawls's Law of Peoples", p. 717.

Practical Reasoning and Transnational Justice

begin from different premises cannot engage with each other constructively, rather than that the result of engagement will inevitably be impassable mutual incomprehension. He denies that it may be possible to *construct standards of reason from a shared morality*; rather the task of political philosophy is to seek out a minimal and limited set of principles on which there is *already* agreement.181

O'Neill does not deny that disagreement is frequently the outcome of intellectual engagement, however she also insists that “there is no reason to be sure that [intellectual engagement]... must fail”.182 That incomprehension and disagreement would be the inevitable outcome of engagement is less convincing if we consider that systems of thought are not *wholly* impenetrable. O'Neill argues that “everyday assumptions” about possibilities of “trade and translation, travel and collaboration... undermine the plausibility of any communitarian conception of practical reason”.183 The following comments from O'Neill draw attention to the porosity of systems of thought, and how our actual engagement with others fails to support the claim that we exist in radically distinct spheres of thought:

If those whose *Sittlichkeit* differ were trapped in mutually impenetrable and incomprehensible discourses abstracting from any mode of discourse would get one no nearer to any other. However, *the actual situation of those whose Sittlichkeit differ is not like this.* Typically, those whose lives are cast in differing terms are able to follow, or to learn how to follow, the terms in which others lives are cast. They may follow and dispute others’ categories and their implications rather than see others’ discourse as beyond the pale of comprehension. The image of radical conceptual isolation depends on an exaggerated picture of differing ways of thought as closed and complete. Ways of thought and life are often (perhaps always) neither. Their boundaries are ill defined; they are porous to and often receptive of elements from disparate ways of life and thought. We are not so much speakers of one of a range of mutually untranslatable languages as we are multilingual. Those who think that modes of Sittlichkeit have determinate and uncrossable boundaries take an idealized view of communities, ideologies, and nations.184 (Emphasis mine)

---

181 This approach has been criticized by Forst as a “contest of modesty”: “[t]aking its lead from Rawlsian concerns about liberal parochialism and the search for political-moral justifications which can be the focus of an international "overlapping consensus," a contest of modesty, so to speak, has developed about the most "minimal" but nevertheless sufficient normative justification for human rights.” Forst, “Justification of Human Rights”, p. 715.


The ways in which persons interact with others across vast distances make it very clear that we can in general translate and follow others’ systems of thought. The “burdens of judgment” only imply that persons begin critical intellectual engagement from reasonable yet disparate perspectives and starting points, which they have arrived at by processes of reasoning. Rawls is not wrong that many considerations relating to judgment suggest that reasonable persons will come to a variety of different conclusions on matters of politics, justice, morality etc., but there is no reason to infer from this that such disparate points of view will persist in being irreconcilable, or at the very least that progress cannot be made toward mutual understanding.

3.2.3. Possibilities for exchange between different traditions on their resources regarding the common good

I noted earlier that a characteristic of Rawlsian citizens and societies is their simultaneously reasonable yet also partially unreasonable nature; while they are capable of endorsing principles of fairness and accepting that they cannot impose their views on others, they are incapable of entering into a fruitful dialogue with others regarding their distinct worldviews and seeking to enlarge their horizons. However, reason implies a capacity to think from and adopt the perspectives of others, and this implies at least the possibility of reaching agreement or compromise on previously divisive matters. As O’Neill noted, there is no guarantee of agreement, but there is also no apparent reason for assuming irreconcilable disagreement to be the unquestionable outcome of debate. Furthermore, this assumption is especially problematic when it guides one’s approach to public reasoning. Taking conflict as a presupposition of an account of public reason at worst rejects and at best obscures the possibility that critical engagement can be productive. Such an approach rules out in advance critical modes of discourse and justifies this exclusion based on unvindicated assumptions.
Practical Reasoning and Transnational Justice

The Jesuit Christian ethicist David Hollenbach also critiques Rawls's account of public reason as an inadequate framework for engagement in the public sphere, and suggests that reasonableness implies a more open attitude toward intellectual engagement and dialogue oriented towards practical challenges. On the limitations of a strategy of avoiding pressing ethical questions in the public sphere, he argues that "[t]he challenges raised by urban poverty and global interdependence... call for a significantly stronger orientation to solidarity than the 'liberalism of fear' and its ethic of tolerance can provide." Like O'Neill, he suggests that our contemporary context is not served by a strategy of disengagement. Avoidance undermines the possibility of mutual understanding and as such is "is unlikely to produce the peaceful, harmonious coexistence it promises" (CGCE, 141). Rather what is required is intellectual engagement, grounded in an attitude of openness to different perspectives, on fundamental questions of the right and the good: "[t]he common pursuit of a shared vision of the good life can be called intellectual solidarity" (CGCE, 137).

He too rejects relativist assumptions about reason and suggests that "[d]ifferences of vision are not so total that we are destined to remain eternally strangers to one another" (CGCE, 138). He notes that while reason is always historically contextualised, this does not imply historical relativism (cf. CGCE, 156). From his particular perspective, he notes that Christian morality is grounded in the assumption of universal access to reasoning about justice and morality: reason derives from God, hence morality should not stand in opposition to reason, rather it must "make sense" (CGCE, 150). Interestingly, he indicates that the virtues of civility that, for Rawls, underpin the capacity of citizens to separate their private and public identities are precisely those virtues that support the possibility of constructive intellectual engagement by promoting respect and cooperation (CGCE, 144).\(^{186}\)

---


\(^{186}\) "Civility manifests citizen's commitment to cooperate with each other in a spirit marked by genuine reciprocity and mutual respect. It is a personal virtue that leads citizens to seek to live together cooperatively". (CGCE, 144)
Hollenbach proposes “dialogic universalism” as the method of intellectual engagement aimed at discovering a shared vision of justice and morality:

Cultural differences are so significant that a shared vision of the common good can only be attained in a historically incremental way through deep encounter and intellectual exchange across traditions. It is also dialogic because it sees engagement with others across boundaries of traditions as itself part of the human good. (CGCE, 153. Emphases mine)

Agreement on shared principles must be the result of a long process of “deep” engagement, and the deliberative process must itself be conceived of as a good. He proposes that deliberative democracy can therefore serve as a way between the apparent dichotomy of, either, the coercive imposition of values, or the avoidance of all potentially divisive ethical and moral issues (cf. CGCE, 141). One crucial insight he offers in this regard is that the human rights and the common good traditions should not be conceived of as mutually exclusive. In his view, human rights are in fact the “guarantees of the most basic requirements of solidarity” (CGCE, 159), one of the key principles of Catholic Social Teaching, for which the “common good” has been the core target. Being formulated as “rights” does not imply that they are undermined by the pursuit of a common ethical vision. On the one hand, without human rights no dialogue is possible, rather a “community of discourse and deliberation can only be so when it is immune from coercion” (CGCE, 164). On the other hand, intellectual engagement between traditions is not a threat to basic human rights because deliberation presupposes respect for the freedom and equality of all participants.

Hollenbach argues that human rights can only be concretised in a dialogue between universal standards and particular contexts:

187 The theologian and ethicist Ethna Regan offers similar insights in her recent book Theology and The Boundary Discourse of Human Rights. She suggests that human rights, rather than themselves constituting the final content of shared standards for morality and justice, function to protect the space for reasoning about the common good. She therefore attributes to rights “the position of protective marginality in ethics: rights are necessarily “marginal” in that they are not ends in themselves, but in this marginal position they play a crucial protective role. Rights do not simply guard the limits below which we should not fall in terms of ethical conduct but are protective of the conditions in which the “more” of ethics—love, virtue, community -- can flourish.” Ethna Regan, Theology and The Boundary Discourse of Human Rights (Washington: Georgetown University Press: 2010), p. 2.
Pursuit of respect for human rights requires an ongoing dialogue about how the universal standards sought by the discourse of human rights relates to the distinctive, particularist self-understandings of the religious communities of the world. (CGCE, 164)

The relevance of abstract rights and principles can only be constituted by their being embedded in particular contexts and traditions. For him this leads to the need to identify shared “goods”; “universal human goods” can only be discovered through dialogue (CGCE, 153). Dialogue involves participants offering visions of the good and *reasons for accepting particular visions*. This claim supports the insight that reasoning with others involves *exchanging reasons for accepting premises*, rather than seeking already shared premises.

In response to Rawls’s assertion that such religious reasons do not meet the criterion of reciprocity necessary for them to count as “reasonable”, Hollenbach points to the long historical process of the formation of ideas of justice and the contributions religious traditions in their discourse and practice have made to the understanding of what constitutes reasonableness and justice, such as the influences of M. Gandhi and Martin Luther King. He concludes that “what Rawls calls the ‘background culture’ plays a formative role in shaping what is politically reasonable” (CGCE, 167).

With regard to Rawls’s reliance on the criterion of reciprocity to ground reasonableness, Hollenbach offers an important insight on how we might interpret such a criterion:

Rawls is ready to accept this influence of religion and culture on the political sphere, but only under the proviso that it be influence under the constraint of reciprocity. If this means that such influence must be exercised *with due respect for the religious freedom of all citizens*, it is fully compatible with the idea of a community of freedom for which we have been arguing. But if it means that we should assume *that existing constitutional democracies and Rawls’s own theory of democracy already know the best way for us to live our common life together*, it must be judged short-sighted. (CGCE, 168. Emphases mine)
If reciprocity can be interpreted as extending respect and recognising the equal freedom of others, this is consistent with a dialogic universalist ethic. However, if it implies that questions of the right and the good, of moral and ethical visions, have been definitively and finally answered by the purely political public reason of a constitutional liberal democracy, it stands in opposition to the requirements of intellectual engagement. Hollenbach concludes that “[t]he task of forming and sustaining a society in which people from diverse religious and cultural traditions can live well together is a never-ending historical project” (CGCE, 169).

What Hollenbach’s account highlights is that the structure of discourse and deliberation, i.e. that it is grounded in principles of mutual respect and equal freedom, is more important with regard to possibilities of agreement than the propositional content of one’s perspective or position. Again I note, in contrast to Rawls’s contradictory description of citizens as reasonable yet unreasonable in discourse, that universal reason implies the possibility for cooperative and productive intellectual engagement. What I would suggest is most crucial in this insight, which is shared with O’Neill, is not that we must be optimistic about the possibility of discovering shared ethical visions, but that we need not be prevented from engaging in discourse through fear of having comprehensive ethical visions imposed on us at the expense of our fundamental rights of freedom and equality; rather fundamental rights are necessary conditions of political discourse.

---

188 I would suggest, as Hollenbach may, that the latter shares more in common with Rawls’s perspective, while it is possible that the former interpretation may be more in accordance with Rainer Forst’s understanding of a “reciprocal and general right to justification”.

189 In this regard, both positions can be compared with Habermas’s argument relating to the shared origins of individual liberty rights and political rights of participation. Habermas argues in contrast to Rawls that deliberative processes need not be conceived of as subordinate to a set of liberty rights that are justified in a consensus that is prior to the deliberative process. Such a priority “demotes the democratic process to an inferior status” (PUR, 128) and “contradicts the republican intuition that popular sovereignty and human rights are nourished by the same root.” Habermas, “Reconciliation Through the Public use of Reason”, p. 129. Pointing to citizens as both authors and addresses of law, he asserts that liberty rights are not endangered by a robust deliberative democratic processes, rather “the dialectical relation between private and public autonomy becomes clear in light of the fact that the status of such democratic citizens equipped with law-making competences can be institutionalized in turn only by means of coercive law. But because this law is directed to persons who could not even assume the status of legal subjects without subjective private rights, the private and public autonomy of citizens mutually presuppose each other... Once the concept of law has been clarified in this way, it becomes clear that the normative substance of basic liberal rights is already contained in the indispensable medium for the legal institutionalization of the public use of reason of sovereign citizens.” Habermas, “Reconciliation Through the Public use of Reason”, p. 130. Emphasis mine.
Like Hollenbach, O'Neill also asserts that there can be no deliberative processes without considering the conditions for sustaining such processes, which in one respect means recognising participants as finite and vulnerable agents who depend on their capacities for agency in order to participate. Also like Hollenbach, she argues that abstract principles risk irrelevance if they cannot be contextualised with regard to the particularities of specified agents and communities. Deliberation on the meaning of principles is essential because implications vary with context.

O'Neill's account does not only support the idea that human rights protect the freedom to participate by rejecting coercion and violence and sustaining agency. It also argues that we must take into account further constraints on discourse in order to prevent it becoming self-defeating. In this way the conditions for communication and discourse O'Neill discusses reflect her interpretation of Kant's account of the vindication of reason; in order for reason to avoid self-defeat it must, aside from ensuring its own freedom, also subscribe to a self-imposed law.

190 O'Neill, "Ethical Reasoning and Ideological Pluralism", p. 720. Specifically with regard to Rawls's apparent use of abstract reasoning, O'Neill suggests that the idealized conceptions it introduces instead undermine its ability to justify itself to various ethical discourses: "[h]ow can abstract reasoning be made accessible or justified to those who inhabit varying modes of Sittlichkeit? Can it be used to vindicate an ideal of mutually independent citizenship? Or is this ideal peculiar to certain political cultures? Is political culture rather than abstract reasoning the secret of the wide accessibility of Rawlsian liberalism in the United States—and perhaps a reason why it has not travelled well beyond the United States?" O'Neill, "Ethical Reasoning and Ideological Pluralism", p. 716.

191 The following passage highlights how O'Neill links the justification of principles of justice (that rejects the search for pre-established agreement) to determinate contexts through an emphasis on deliberation: a "plurality of human beings between whom there were deep differences of outlook could no more presuppose shared ideals of citizenship to guide them to principles of justice than they could appeal to preestablished harmony. Rawls relies on the separation of the right from the good to underpin his appeal to selective sharing of ideals against a background ethical pluralism. As we have seen, this separation and the corollary subordination of Sittlichkeit to Moralität is a central criticism made against deontological liberalism. Liberals who reject Rawls's move must look for achievable rather than presupposed agreement. They can begin by asking what can be done by those who find themselves confronted with the pluralism that is constitutive of modern circumstances of justice, yet seek to vindicate universal principles. They cannot (without rejecting the search for universal principles) appeal to some specific ideal of citizenship or to other non-universal features of human choosers: hence they must seek abstract but nonidealized principles of justice. However, if their reasoning is not to be inaccessible to many audiences, they must also find ways of deliberating that connect these abstract principles of justice to the more specific and accessible categories of discourse of particular communities. This means that liberal reasoning that both answers the critics of deontological liberalism and retains internationalist commitments would have to address the charge of abstraction in both of these ways. O'Neill, "Ethical Reasoning and Ideological Pluralism", p. 717-8.
O'Neill notes that the “first task” of reasoning beyond the borders of distinct systems of thought “must be to enable communication”.\textsuperscript{192} She argues that Kant’s own account of public reason can provide the framework for a discursive process of reasoning that can construct shared standards for reasoning: that it “offers a way between the cliffs of transcendent vindication of reason and the whirlpools of relativism” (PUR, 538).

3.2.4. Tolerance as a necessary condition of Kantian public reason

Like Rawls, O’Neill’s Kantian account of public reason also begins from the modern context of justice: ethical pluralism. However this account understands ethical pluralism as a plurality of \textit{uncoordinated} agents who do not share premises on which to ground further agreement.\textsuperscript{193} Rawls’s approach fails to offer an account of reasoning suitable to this context because it restricts the audience to whom justification is offered. Ethical pluralism demands a form of reason that is not restricted to a particular group that share particular premises. Public reason in this context must be reason that does not derive its authority from heteronomous (and private) sources of authority. In this way we begin to see that Kantian public reason is grounded in the \textit{vindication} of reason, which was discussed when we examined O’Neill’s Kantian foundations. Reasoning that relies on restricted premises, such as a shared agreement, \textit{cannot survive scrutiny} in the \textit{actual} public sphere and in this way cannot vindicate its own authority:

Kant sets out a dilemma... if we try to constrain or control attempts to reason - we lose the very justifications we seek. Discourse that defers to authorities that lack reasoned vindication achieves at best restricted scope and authority, and those who buttress their conclusions by appealing to

\textsuperscript{192} O’Neill, “Abstraction, Idealization and Ideology in Ethics”, p. 68.

\textsuperscript{193} “The modern circumstances of justice are ethical diversity: people have variously constituted capacities to reason and to act and varying forms and degrees of independence from one another and from the institutions and practices that constitute their \textit{Sittlichkeit}”. O’Neill, “Ethical Reasoning and Ideological Pluralism”, p. 718.
authorities they do not vindicate end up relying on the dubious merits of an argument from authority.\textsuperscript{194}

In this way, constraining or restricting attempts at reason undermines justification, i.e. it suppresses the criticism necessary to test the shareability of principles for reason or action. Where reasoning is restricted and criticism suppressed or avoided, deficiencies will be concealed. It is for this reason that Kant asserts that "the public use of reason should always be free" (PUR, 525). Kant suggested that only an inclusive approach to public reason could hope to discover shareable and hence vindicated principles or standards of reason:

In this way the powers and shortcomings of reason can best be revealed, its authority delimited, and antinomies avoided. Reason's authority consists simply in the fact that the principles we come to think of as principles of reason are the ones that are neither self-stultifying nor self-defeating in use. The best way to find which principles have this character is by encouraging the increasingly public use of reason. Indeed, if reason has no transcendent foundation, there is nothing else that we can do. (PUR, 534-5)

Given that there can be no transcendent foundation of autonomous reason, testing reasons in free exchange is the only way to discover whether they can function as shared reasons. Rawls was also interested in reasons that could be made "public", however we have seen that his focus was ultimately on what was likely to be agreed on, which caused him to restrict the plurality and to therefore introduce contingencies that in fact undermine the actual public character of reason. In contrast to Rawls, "Kant takes it that the capacity of principles and standards to be made public to all others without restriction is constitutive of their reasonableness" (PLPR, 14. Emphasis mine). Reasoning that cannot withstand criticism from a fully public perspective (for example, if it is grounded in an appeal to an argument from authority) cannot be considered fully reasoned.

For standards of shared reasoning to have any authority they must be public. This implies that processes of reasoning must be free and open. As we noted earlier, presupposing the contingencies of a particular society is not a promising strategy

\textsuperscript{194} O'Neill, "Bounded and Cosmopolitan Justice", p. 53.
in the context of porous boundaries and conflicting identities. “Private” reasoning, in the sense of presupposing institutional roles, cannot have authority in this context; rather “to accept and foster the authority of reason is to submit disputes to free and critical debate” (PUR, 535). The obvious implication of such considerations is that, for processes of reasoning to be free and critical, we must be receptive to others’ communication; we must not dismiss, avoid or reject their communications. It is for this reason, according to O’Neill, that Kant argued for the necessity of toleration in critical discourse and communication. Toleration must be a “precondition” of the process of seeking shared standards of reasoning (PUR, 535). Rather than arguing for an “instrumental justification” of toleration,

Kant’s thought is rather that a degree of toleration must characterize ways of life in which presumed standards of reason and truth can be challenged and so acquire the only sort of vindication of which they are susceptible. (PUR, 535. Emphasis mine)

What is implied by toleration is that criticism is heard and responded to; without toleration critical engagement is not possible, and hence the authority of reason cannot be constituted. Toleration as the guarantor of free and open debate is crucial to vindicating the authority of reason and discovering shared standards of reasoning.

Kant’s conception of toleration as set out by O’Neill is an active and demanding conception. It is also markedly different from Rawls’s conception of toleration. Toleration, for Rawls, both at the domestic and international level of public reason, amounts to a strategy of non-engagement or avoidance on matters of disagreement. To avoid the inevitability of conflict on contentious issues, Rawls argued that they must be removed from the sphere of public debate and, crucially, replaced with pre-established agreement. Toleration on this account is a form of “acceptance” of alternative accounts of reason only on the proviso that they remain within a “private” sphere where they do not impose themselves on other perspectives. In Part 2, I argued that, with regard to international justice, the

---

195 O’Neill quotes Kant on the grounding of reason’s authority in the ability to withstand reason-based criticism: “Criticism and the toleration that criticism requires are fundamental for the authority of reason, and we are recommended to ‘allow your opponent to speak in the name of reason only and combat him only with weapons of reason’”. (PUR, 534)
apparent consequence of this conception of tolerance was the “acceptance” of, or acquiescence in the face of, unjust political institutions. O’Neill has argued that such restrictions, along with a reliance upon shared understandings and identity, are ill-suited to the actual international context where domestic boundaries cannot be presumed to circumscribe homogenous peoples. On a more fundamental Kantian level, non-engagement, as the avoidance of criticism, implies that Rawls’s account of public reason has no claim to authority. As I argued in Part 2, it is therefore not compatible with an objective public perspective; the assertion of objectivity via consensus or agreement lacks authority.

Toleration on Rawls’s account may not even constitute tolerance. According to O’Neill, toleration as avoidance or non-engagement seems to be “too negative a matter to be fundamental” (PUR, 526). Only if one could view the object of toleration as “self-regarding” expressive acts, could non-engagement or avoidance possibly be considered sufficient. However, O’Neill notes that “the standard point of expression is communication” (PUR, 526). Expressions, in general, seek to reach an audience, to be heard and to be understood, i.e. they seek to communicate. Non-engagement is a deficient conception of toleration where the object of toleration is communication:

All successful communication requires some sort of recognition or uptake by others, whether this consists in an understanding of the context communicated or merely in recognition that the other seeks to communicate... we do not tolerate others’ communications if we are merely passive and noninterfering. (PUR, 527)

Communication is an intersubjective activity; it is a way for agents to engage with other agents. It involves imparting, transmitting and exchanging information; it is a form of contact and connection. On this understanding, passivity is an unsuitable strategy for responding to what is communicated. We cannot equate “doing nothing” with having “no effect”, rather non-engagement “may convey disapproval or hostility”, and if not, may at least “indicate that any communication is a trivial and indifferent matter, not worthy of discussion or refutation, a merely private affair” (PUR, 526).
In her article, “Practices of Toleration”, O’Neill points out that we must assume that “democracy requires the possibility of public communication” and that it is therefore inadequate to view the object of toleration as “self-regarding” forms of expression.\footnote{O’Neill, “Practices of Toleration”, in Judith Lichtenberg ed., Democracy and the Mass Media (Cambridge: Cambridge University Press, 1990), pp. 155-185, p. 170.} For Rawls too, public communication is essential to a functioning democracy. However, for Rawls, public communication is not the object of toleration. Public reason on his account is grounded in what is already agreed upon and there is therefore no need for toleration of such communications. A need for toleration would imply that what is expressed is in opposition to one’s own position, and because areas of disagreement are removed from the agenda, toleration is unnecessary. However, I have argued, based on O’Neill’s Kantian interpretation, that public reason or public discourse cannot ground itself in what is already agreed upon. It must then be assumed that what is communicated in the public sphere represents divergent and possibly opposing perspectives on basic political matters, which are communicated in the expectation of a response. Tolerance must therefore apply to public communication.

Seen in this light, toleration is a demanding conception. It is a matter of listening to those who disagree with us and attempting to follow their reasoning and respond to them. It is not a matter of restricting the context and content of public reason; such an approach cannot discover shareable standards in the context of real ethical plurality, and correspondingly it cannot answer the questions of basic justice that arise in this context. Rawls’s conception of toleration as non-engagement is wholly inadequate where there are fundamental disagreements concerning matters of political identity and the justification of principles of justice. Rather, toleration is a necessary condition for public reasoning; it ensures that reasons will not be a priori excluded from the public sphere and guarantees the free and inclusive public reason without which no vindicated, shareable and universal standards of reason can emerge.
3.2.5. Constraints on public reason as enabling conditions for communication

Recall that the vindication of reason implies that reason can follow no law other than that which it gives itself. Free reasoning prevents domination by external authorities; public reason must follow no external law. However, O’Neill also stresses that reason that follows no other law cannot simply follow no law. Lawless reasoning can only lead to incomprehensibility:

A public use of reason, we have seen, is in the first place one which could reach the world at large if suitably publicized. It must therefore assume no authority which could not be accepted by an unrestricted audience. Since “the world at large” accepts no common external authority, the only authority the communication can assume must be internal to the communication. (It cannot on Kant’s account, and on many others, assume no authority whatsoever: “lawless” communication ends in gibberish and loss of freedom to think). The only authority internal to communication is, on Kant’s view, reason. (PUR, 531)

There is no common form of private reasoning that can reach the “world at large”, and therefore the authority of reason must be internal. This implies that more than a rejection of external authority is necessary to ground the possibility of discovering shared standards in reasoning. Although freedom prevents domination, radically free discourse will not have the resources to resist domination by external authority, and risk devolving into agreements that do not meet conditions for justifiability. Lawless reasoning is self-defeating and therefore reason and communication must be structured and constrained by certain self-imposed conditions. While restrictions such as those imposed by Rawls prevent and disable communication between ethical perspectives, other restrictions may enable communication.

O’Neill suggests that if we conceive of reasoning as a practical problem we may gain insight into what necessary constraints must be placed on our reasoning. Reviewing her arguments in this regard will highlight the derivation of Kant’s account of public reason from fundamental considerations about agency, reason and autonomy. O’Neill comments:
Because the structure of human communication is not preestablished, its conduct is a practical problem. We are not guaranteed coordination with others, so must ask which maxims or practical principles can best guide us when we seek to communicate, and must try and avoid principles which could not regulate communication among a plurality of separate, free, and potentially reasoning beings. If we find such “principles of communication”, their justification must be recursive; they will simply be principles by which practices of communication can be maintained and developed rather than stultified. (PUR, 540)

The justification of standards and principles for public reasoning is both recursive, which in this context implies it must be established by application in discourse, and modal, which means it is given by rejecting principles that undermine possibilities for action, reason and communication. These two features of grounding the authority of public reason may appear to be incompatible if one considers them in terms of the tension O’Neill’s account highlights between the actual agreements established in discourse and justifiable agreements that meet certain conditions. O’Neill suggests that she “shall try to show that Kant offers the appropriate complement to a discursive grounding of reason in his reasoned grounding of practices of discourse. Reason’s authority and toleration are interdependent” (PUR, 540). Rather than suggesting one or the other is entirely fundamental, she suggests that the recursive testing of the shareability of reasons in discourse, and the constraining of discourse by principles for reasoning and action, are mutually supportive.

The insight of Kant’s supreme principle of practical reason is that the only available vindication for practical reason is that its principles are “capable of guiding the interactions, including the communicating, of beings whose coordination is not naturally guaranteed” (PUR, 541):

This is quite explicit in the Formula of the Kingdom of Ends and not far below the surface in the other formulations. The idea of acting on maxims fit to be universal law, which is the core of the Formula of Universal Law, invokes the notion of a plurality of free and rational agents who act only in ways that do not preclude others’ doing likewise. The idea of treating all others as ends, which is the core of the Formula of the End in Itself, invokes the notion of a plurality of agents who control their action to achieve coordinated respect for one another’s freedom and rationality. These
Practical Reasoning and Transnational Justice

standards can be applied reflexively to the process in which they are established. They must be applied reflexively if they lack transcendent vindication and are yet to have the authority of principles of reason. (PUR 541)

As we saw when we examined O’Neill’s Kantian foundations, the formulations of the Categorical Imperative are distinct yet equivalent formulations of the basic principles that must guide the reasoning and action of a plurality of finite and vulnerable agents who lack pre-established agreement or shared premises for reasoning. “Must” is understood as a modal concern implying conditions that are necessary to sustain a plurality of agents. Therefore it implies rejecting principles that undermine capacities for agency, such as violence, coercion and deception. Reasons or maxims based on such principles are self-defeating: reasoning can only take place between at least partially rational agents, hence destroying capacities for agency destroys the very context for reasoning. These considerations shape the structure of Kant’s account of public reason and communication:

Incipiently free and rational beings, who lack transcendent principles of practical reasoning, can and must regulate their communicating by maxims that do not undermine or stultify their incipient communication. There is nothing else they can do if their communicating is neither transcendentally nor (fully) naturally coordinated. In its application to maxims of communication, as to other maxims, the Categorical Imperative is no more than the test whether what is proposed is action on a maxim that could be shared (not “is shared” or “would be shared”) by a plurality of at least partially free and rational beings. (PUR, 541)

Reasons that can be shared are those that preserve the context for reasoning and that support and respect the agency of others. At the most basic level, maxims of communication must be shaped by considerations of what sorts of actions destroy or undermine communication.

O’Neill quotes Kant that “reason has no dictatorial authority, its verdict is always simply the agreement of free citizens” (PUR, 533). However, again we must note that actual agreement alone is not sufficient or fundamental; rather those who seek agreement must also consider the conditions that can make agreement morally justifiable:
Reason, on this account, has no transcendent foundation but is rather based on agreement of a certain sort. Mere agreement, were it possible, would not have any authority – what makes agreement of a certain sort authoritative is that it is agreement based on principles that meet their own criticism. In Kant’s view such self-criticism is best sustained in the form of free, critical and universal debate. (PUR, 534)

If freedom and openness are necessary to facilitate reasoned communication in an ethically diverse context, it follows that the conditions that make free and open communication possible must be sustained. Kantian public reason depends on modal considerations of possible action, communication, reason and agreement:

As in Rawls’s account, so in Kant’s, the constraints of public reason do not derive from metaphysical sources, from individual motivation, or from formal requirements. However, in contrast to the Rawlsian account, Kant does not derive an account of public reason from an assumed shared sense of political identity. Kantian public reason is no more than the modal requirement that underlying principles and standards of thought, communication, and action be accessible to any audience, so be something to which they can agree, even when it is not something to which they will agree.¹⁹⁷ (PLPR, 18)

Both conditions for free agency and the testing of reason in an open public forum are necessary for vindicating the authority of a reasoned agreement. This is how a “discursive grounding of reason” is complemented by the “reasoned grounding of practices of discourse”. As Hollenbach notes, toleration prevents the a priori exclusion of reasons from the public sphere, and rights (or institutionalised principles of justice) constrain and thereby enable discourse by sustaining the capacities for agency that make participation possible. The institutionalisation of robust and inclusive, context-sensitive, discursive practices must coincide with the institutionalisation of principles of justice, in order to prevent discourse being undermined by practices that undermine agency.

¹⁹⁷ As we saw in Part 2, O’Neill’s Kantian account takes the conditions for agreement as prior to actual agreement and holds that without such conditions justification is in question: “[t]he justifications offered by a Kantian conception of public reason will not be contractualist. They are mute on the question of whether an actual or even hypothetical convergence of wills on any specific standards of principles is likely or whether if it emerged it would amount to any sort of justification. It allows that a consensus can be iniquitous.” (PLPR, 18-19. Emphasis mine)
However, enshrining principles of justice that reject maxims which undermine human agency may not be sufficient: we must also ask which other practices of protecting and regulating communication may be important to prevent the emergence and domination of ideologies that (even without violence or threat to communicators) exclude or silence.\(^{198}\)

Tolerance and inclusivity demand more than rejecting coercion, violence or deception:

Practices of toleration demand not merely that communicators neither coerce nor destroy one another, but more specifically that they act to institute and sustain practices of communication that exclude nobody. (PT, 167)

Tolerance involves sustaining communication, and this implies more than rejecting principles that undermine agency. For this reason, O’Neill argues that a rights-based perspective on tolerance and communication in the public sphere may be inadequate.\(^{199}\) Such approaches focus only on how we ought to treat communicators, i.e. rights of free speech etc., and not on how communicators should act. As with our earlier analysis of the obligatory yet selective nature of imperfect obligations, it is evident that rights-based accounts fails to clarify “what can make a rights-respecting communicative or expressive act wrong” (PT, 156). That obligations have no corresponding rights and are therefore not “claimable” does not imply that they are not obligatory. Rights may be respected by acts or practices that undermine the conditions for sustaining practices of communication; where this is the case respect for rights alone cannot protect public reason. There is good reason to consider the ethics of communication then from the perspective of obligation. As we have seen, O’Neill argues that the justification of rights are best derived from thinking about obligation, about principles for action, and that “thought about rights constitutes a proper part of

\(^{198}\) O’Neill, “Practices of Toleration”, pp. 175-6. Further references in text (PT)

\(^{199}\) From “Practices of Toleration”: “I shall dispute the coherency and adequacy of liberalism that is agnostic about the good of man, by sketching some of its implications for the ethics of communication”. (PT, 156)
thought about obligation” (PT, 157). In the particular case of the ethics of communication, rights-based perspectives obscure the harm that is caused by action that respects rights whilst undermining discourse.

The necessary conditions for dialogue and discourse cannot be encompassed by accounts that believe they have discovered the “largest liberty compatible with like liberty for all” – especially if they also hold that a “liberty of the person” “fills’ any ‘space’ remaining when other rights have been specified”, as this liberty is not consistent with imperfect obligations (PT, 162). An ethics of communication must therefore be orientated toward both the rights and the obligations of participant communicators.

3.2.6. Beyond freedom of expression: communicative obligations

As we have seen, if we are to regulate our communication by maxims that do not undermine it, we must reject maxims grounded in violations of the capacities for agency that underpin communication. O’Neill suggests that on Kant’s account then forms of reasoning that rely on either polemical or deceptive forms of communication must be rejected. Polemic seeks compliance rather than comprehension and agreement, and deception too is not comprehension (cf. PUR, 541-2). O’Neill continues that “[r]estraints of polemic and lying may be necessary guidelines for tolerating one another’s communicating, but they are only the beginning of practices of toleration” (PUR, 542). Toleration is necessary so that reason can be free from external and unvindicated sources of authority. Where practices of communication are dysfunctional, communication is still at risk of being undermined or dominated by ideologies that lack authority. O’Neill notes that “[w]e have plenty of evidence of the power of established and establishment modes of thought, and the ways in which these are used to set the terms of debate, to co-opt criticism and to marginalize dissent” (PT, 173. Emphasis mine).

200 Again O’Neill does not deny the importance of rights, rather she asserts only that “[a]ll beliefs about rights can be expressed in terms of correlative (perfect) obligations”. (PT, 161)
As an example, she warns against the encroachment of "Newspeak" into the sphere of public reason, i.e. of the homogenising of the language of public reason. Newspeak, by removing shades of meaning from language, limits the resources for constructing arguments and reasons, and therefore undercuts the diversity of language that is necessary to vindicate reason and discover shareable principles. It misleads by imposing oversimplified terminology and silences by inhibiting languages' ability to express dissent. By inhibiting opposition, such restricted forms of language also threaten the fundamental conditions of democratic society (cf. PT, 174). O'Neill argues that rights-based approaches cannot account for why Newspeak is incompatible with democratic public reason. Provided a mode of communication does not violate rights, it is acceptable from a rights-focused perspective, however, O'Neill notes that modes of expression can seriously damage communication without violating rights and hence it is important to ask whether mass communication even in democratic societies may establish categories and modes of discourse that so dominate discussion and thought that they endanger the very formulation and communication of opposition... Wherever people find it hard to formulate or discuss the thoughts of opposition (whether antisocialist or anticapitalist, anti-Islamic or antinationalist), we may suspect that established ideology whether or not imposed by overt exercises of state power threatens democratic communication. (PT, 175)

Public reason depends on the free exchange of reasons; where some perspectives cannot even formulate their claims or reasons within the framework of the public sphere then the capacity of discourse to construct shared standards of reasoning and justified agreements is endangered. O'Neill asserts therefore that, beyond respect for rights, an account of practices of toleration needs to look at the wider context of communication and at acts and practices that may exclude or silence while still respecting rights.

Where free and constructive communication is endangered or subverted by ideology or established categories, a state policy of noninterference in discourse will effectively erode public reason and communication. According to O'Neill, laissez-faire policies "merely assign the regulation of communication to nonstate powers (PT, 178). What this implies is that, where the state fails to institutionalise
ways of sustaining modes of discourse, free and open discourse cannot be sustained and risks being hijacked by other systems of thought that do not seek shared standards for reasoning. In this way drawing a distinction between political/legal regulation as “interference” and commercial/educational/media regulation as “noninterference” is misleading and dangerous (PT, 178). O’Neill suggests that a “better and less abstract aim [than a focus on freedom of expression] for a democratic society is a set of practices that enables a wide range of communication, especially of public communication, for all” (PT, 178).

An important consideration for an account of the ethics of communication is that reason is interactive: O’Neill quotes Kant as asking “how much and how correctly would we think if we did not think as it were in community with others to whom we communicate our thoughts”. She continues:

The standards of reason cannot be found in solitary thinking: on the contrary those who seek to reason must structure their thought, speech and communication in ways that others can follow. This double modal constraint is fundamental to the positive aspect of Kant’s account of public reason.

On this basis Kant offers several insights into the maxims or obligations that those who seek to communicate must adopt and adhere to. He suggests three maxims that must guide our communication if we are to sustain communicative practices and avoid mutual incomprehensibility. The first was that agents or reasoners must think for themselves; given that communication takes place between agents, total failure to preserve a measure of separateness from those with whom one supposedly communicates is self-defeating (PUR, 543). This maxim includes the requirement to apply the universal principle of the use of reason, in order to avoid the imposition of alien authority on one’s reason. Therefore Kant also referred to it as the “maxim of the self-preservation of reason” (PUR, 543). Where would-be reasoners fail to think for themselves the context for toleration is eliminated: where there exists “no plurality of viewpoints”, “toleration becomes pointless” (PUR, 543). The second maxim urges taking the standpoint of the other: the

---

Practical Reasoning and Transnational Justice

"maxim of enlarged thought" (PUR, 544). Again, note that reason is directed toward others and that communication takes place between agents. Only reasoning which can adopt the perspective of the other can be accessible in principle to the audience it seeks to reach; reasoning which is incapable of thinking from the standpoint of another amounts to a mere expression of one's own perspective, and hence no communication has taken place (PUR, 544). The third maxim asks for consistency in thinking; given that discourse and debate can demand a constant revision of perspectives and judgments, and because human reason is not "antecedently given" but must be "gradually developed", the demand that the agent attempt to maintain consistency of thought is not trivial (PUR, 544).

O'Neill also suggests further implications of communicative obligations, based on the active nature of toleration. If free and open discourse is a necessary feature of public reason then toleration must involve taking others' perspective seriously. This can have demanding implications:

[A] deep commitment to respecting and hearing other voices must always take seriously the thought that other voices have been smothered and silenced, so that mere opportunity to speak may not enable them to become audible. (PT, 173)

On an institutional level, this implies that democratic societies must sustain institutions that "foster diversity of communication and protect positions and voices that are in danger of being silenced or marginalized" (PT, 173). Where diversity in thinking is not actively supported there is a great risk that it will be eroded by dominant systems of thought or ideologies. O'Neill argues that communicators who aim to reason publically are obligated to conduct their communication in accordance with certain conditions: communicators must aim to be interpretable by all, they must "assume and reveal their own fallibility" and, correspondingly, they must "stress rather than suppress the difficulties of their own viewpoint" (PT, 173. Emphasis mine). O'Neill acknowledges that these conditions are not easy to follow, and that there "may be many reasons for shunning the hard and often boring task of writing, speaking, or editing for a heterogeneous audience in a way that is open to challenge and shows that and how it is open" (PT, 172). In
the context of modern mediated communication a “large range of techniques can
disguise communicator’s fallibility and their latent agenda” (PT, 172). It is all too
tempting and easy to shape seemingly straightforward communication toward
particular agendas, for example through the use of loaded terminology, or to stress
or obscure aspects of one’s arguments depending on their relative strength or
weakness.

Another damaging practice, often perpetuated by media outlets (who may
themselves by unaware of it), is the “neglect of views that lie beyond the assumed
limits of ‘acceptable’ opinion” (PT, 172). Diversity of perspective is at risk where
the selectivity of mediated communication is unregulated. Overall, O’Neill
concludes that there are a great many methods that “can be used to sway public
opinion while disabling public criticism” (PT, 172). A free space for
communication is not an un-manipulated space; rather a radically free space is
deeply vulnerable to manipulation by agendas that are potentially incompatible
with the public good. A nuanced account of communicative practices would need
to consider the more opaque and indirect practices that can damage and erode
possibilities for communication and discourse between free agents.

O’Neill also notes that the maxims of communication “[c]annot be fully specified in
the abstract” (PT, 168). Like principles of justice, they can only be specified by
taking account of the specific contexts to which they are to be applied, and “no
unique set of policies” will fit all contexts (PT, 178). There may be many aspects of
distinct contexts that are relevant to communicative ethics. For example, O’Neill
notes that “[w]hen communication is mediated the requirements for
interpretation are again transformed” (PT, 169). Direct and mediated
communication may imply distinct types of more specific communicative
obligations, depending on which acts and practices can be identified as
endangering or suppressing communication. O’Neill concludes therefore that
“[w]e cannot know what it takes to preserve and sustain the presuppositions of
communication until we know what they are in a particular context” (PT, 169). I
would add that given that the effects of particular policies and regulations are
often opaque, their efficacy at sustaining communication and discourse can only
be evaluated in context. However, despite the unavoidable non-specificity of an abstract ethics of communication, it does offer at least a "sketch" of principles of communication, and crucially it shows why maxims of communicative demand more than respect for others' rights (cf. PT, 169).

The Categorical Imperative is the consideration that underpins the ethics and maxims of communication in O'Neill's account. That principles and reasons be shareable is a modal concern regarding the possibility of communicating and acting in pluralist contexts. In this way the maxims of communication are grounded in the principle of toleration that guarantees the freedom and inclusivity necessary to test the shareability of principles or reasons. Without an active and regulative approach to toleration, discourse is at risk of being undermined by dysfunctional forms of communication. It is also at risk of being self-defeating where communicators do not take the voices and perspectives of others seriously. Toleration is necessary so that dissenting voices have a space to object when they cannot share a maxim or a proposed course of action. Furthermore, it suggests that there is little reason to tolerate modes of communication that impair public uses of reason. O'Neill summarises the grounding of communicative obligations thus:

If there is a possible form of communication between beings who are separate and whose coordination is not naturally given or preestablished, then those beings must guide their attempts at communicating by principles which neither erode their own thinking nor fail to seek to understand and to follow the thinking of others, nor shrink from the task of working through and integrating a consistently revised set of judgments to achieve consistency. (PUR, 545)

Without these constraints, shared standards of reason cannot emerge in the context of ethical plurality. In the same way that obligations of justice enable

203 On the textual basis for O'Neill's interpretation of tolerance and communication in relation to Kant's foundational formulations, she comments: "confirmation for this reading of the authority of the Categorical Imperative and its close connection with practices of toleration can be found in Kant's comments on communication. Although the notorious four examples of the applications of the Categorical Imperative of the Groundwork do not include any specific maxim of communication (some false promising may be a failure in communicating), in other works Kant says a good deal both about unacceptable and about morally required maxims for communication". (PUR, 541)
rather than hinder capacities of agency, constraints support greater freedom in communication: "of course, all regulation restricts; but in a world of interdependent and mutually vulnerable agents it also enables" (PT, 180). Hence the task of concretising practices of communication involves attempting to discover and protect the practices that "best enable and least obstruct communication here and now" (PT, 180-1). Deliberative democratic processes will be futile without meeting conditions for effective and successful communication.

3.2.7. Seeking consensus in pluralist contexts: the grounding of deliberative toleration in moral justification

In pluralistic democratic societies the demands of public reasoning are not easy. Treating others’ perspectives and styles of communication with respect means following “unfamiliar categories of discourse” and not imposing “established categories of discourse on those who do not share them” (PT, 174). In order to take diversity seriously, practices of communication must be shaped with an understanding of the many ways minority and dissenting voices can be marginalised and silenced. Without taking cognizance of these dangers, and recognising that communicative obligations exist, communication is in danger of failing to be even comprehensible let alone reaching agreement on standards of reasoning and questions of basic justice.

The political philosopher James Bohman also critiques Rawls's account of public reason as incompatible with contemporary contexts of “deep” cultural pluralism. Like O’Neill, he argues that the porosity of boundaries and complex interactions between conflicting identities present great challenges to the public reason of liberal democratic societies. Also like O’Neill, he suggests that public reason cannot be defined in accordance with the shared ideals of liberal citizens. He questions

---

204 Bohman comments: "The emerging challenges to the liberal regime of toleration even in its expanded multicultural form are increasingly transnational, given the fact that global migration has spurred new levels of pluralism in liberal democratic societies. This migration will call into question the requirements of citizenship, as people no longer live their lives within the boundaries of one nation-state... Given the fact of deep pluralism, cosmopolitanism now begins at home. It may well be that the deliberative framework in societies characterized by migration and deep pluralism
the suitability of reciprocity as a condition of reasonableness in democracies with multiple constituencies and subcultures, noting that in such contexts "what higher order principles such as fairness consist of may be precisely what is at stake".\(^{205}\)

Reciprocity cannot function as a guiding principle where it is the equal status of participants that is in question, i.e. where the terms of the debate themselves are in question.\(^{206}\) Bohman’s account draws attention to the inadequacy of a conception of reasonableness that relies on shared liberal values to ground its objectivity; he asks “[w]hat role does reason play in such deliberations if standards of rationality are themselves subject to deeply conflicting interpretations?”\(^{207}\)

Bohman also offers the insight that such standards equate those who are actually unreasonable with democratic groups that “want to extend or modify reasonable political consensus”.\(^{208}\)

Bohman argues that, if citizens seek \textit{mutually acceptable} solutions to conflicts, deliberation means more than “silent toleration”, and therefore non-engagement is too minimal a basis for public reason (DT, 758). He argues further that non-interference by a majority may also be \textit{undemocratic} (cf. DT, 758). Rawlsian toleration is inadequate as “it is unlikely citizens will be able to deliberate about the sources of conflict” (DT, 762).

Like O’Neill, he asserts that toleration is “at minimum discursive openness” (DT, 763). He grounds this with the claim that citizens’ equality is “manifested in
sustaining conditions for communication” (DT, 763). On this account then “toleration in a deliberative democracy is based on the commitment to the principle of political egalitarianism: that is, the equal availability of or access to political influence for all citizens over all decisions that affect them” (DT, 769). Toleration and inclusivity in deliberation derive from the equal status of citizens of a democratic society, not all of whom share liberal ideals.

Deliberative toleration must not then proceed by falsely generalizing the perspective of the dominant group (cf. DT, 765), for example by constraining public reason in accordance with a particular shared set of liberal ideals. In contrast to Rawls, Bohman argues that public reason is not about the “functional requirements for stability” but about *respecting the voices* of those who dissent (DT, 766). In the case of religious objections, he suggests that they could be legitimate where they contest “certain regulative principles that guide deliberation and its regime of tolerance” (DT, 767). Rawlsian public reason is in opposition to democratic commitments: “[n]ot to offer justification even to the unreasonable is to exclude them from the community of judgment and thus to violate the democratic commitments to political egalitarianism and nondomination” (DT, 768). That citizens appear unreasonable is not sufficient to deny them justification, rather their apparent unreasonableness “does not undermine their reflexive challenge” to the regime of toleration and public reason (DT, 768. Emphasis mine). He suggests that the purpose of toleration is to protect communication where disagreement threatens the deliberative process (cf. DT, 774). He stresses that the standards that constrain deliberation must be democratic; they must promote “free and open communication” and make possible “effective participation of all” (DT, 770). In this way he contrasts a *deliberative democratic* interpretation of norms as *enabling conditions* with a *liberal* interpretation of norms as *constraints*. Reflexive challenges must still be grounded in democratic ideals, either by appealing to freedom and equality or by participating in “providing a new perspective in the process of dynamic reflective equilibrium concerning democratic ideals” (DT, 774).
What Bohman highlights is that in the actual context of pluralist democracies there is a tension between democratic ideals and the principles of liberalism. Equality and reciprocity may be at odds where it is the status of citizens that is at stake, and realising equality may imply accommodation of distinct ideals, rather than an expectation of substantively shared ideals in the public sphere. He suggests that "mutual criticism and interpretation" must then replace standards of Kantian "impartiality" in the public sphere. While it is not possible here to critique his understanding of impartiality and how it applies to various "Kantian" positions on public reason, it is likely that much of his objection is shared by positions such as O'Neill's, which rejects both de-contextualised and idealised interpretations of abstract Kantian principles.

What is also of interest is that Bohman appears to derive the need for a free and inclusive form of public reason from the challenge of living up to the ideals of freedom and equality in contexts of real pluralism, contexts where specifically liberal values cannot be presumed to be shared by all citizens. On his account then it appears that the justification for standards of communication is separate from and derivative of the justification of principles of freedom and equality. When compared to O'Neill, this highlights a distinctive and pivotal aspect of her account: she presents democratic ideals and public reason as sharing the same fundamental justification. O'Neill traces the inclusive and diverse nature of public reason back to the Categorical Imperative and the supreme principle of practical reason, and locates the constraints on public reason in the law-giving capacity of autonomous moral reasoning. Public reasons are constrained by the same considerations that prevent moral reasons from being shareable, and these considerations relate fundamentally to respect for the agency of other persons. In this way her account can justify an open and universally accessible form of public

---

209 Bohman comments: "[w]hat is interesting about this procedure of compromise formation is that it does not require any kind of abstraction to an impersonal standpoint; each party can modify the text according to their own values and principles, and the result does not necessarily reduce the plurality of points of view in the same way as impartiality implies. The text is precisely the ongoing framework of deliberation, which each time must be modified enough to assure continued cooperation by both sides. Bohman, "Public Reason and Cultural Pluralism", p. 270. And: "[p]ublic reason works to transform the cultural framework of each culture through mutual criticism and interpretation". Bohman, "Public Reason and Cultural Pluralism", p. 272.
reason while simultaneously justifying the fundamental moral obligation from which human rights and democratic ideals are derived.

Both Bohman and O’Neill reject Rawls’s suggestion that a set of substantive constraints must restrict public reasoning. However, O’Neill argues that only morally justifiable principles can ground the possibility of sharing reasons, and these must therefore constrain discourse. Such constraints then involve sustaining free agency, free discourse and rejecting dysfunctional modes of discourse. Yet, as we have noted, she emphasises that the relationship between the constraints on discourse and discourse itself is not unilateral or asymmetrical. While abstract obligations for protecting reasoning and communication constrain discourse, shared standards of reasoning also emerge from discourse as those standards that have been accepted as shareable by a diversity of perspectives. In this way

[the grounding of principles of reasoning in incipient communication is mirrored by the grounding of developing communication in principles of reasoning. The supreme principle of practical reason both emerges from and disciplines human communication. (PUR, 545)]

What is actually shared can only be discovered in discourse. However, given that actual agreement alone is insufficient for moral justifiability, discourse must also constrain itself based also on considerations of how reason and action can be undermined. What is shareable must be shareable both in the modal “possible” sense and in the sense of being actually shared. Respecting agency remains the fundamental guiding constraint on action (now in the sense of communication) yet, as we have noted, such guidelines must be contextualised in order to determine their relevance to particular contexts, traditions and self-understandings.

In this way Kantian constraints on public discourse do not offer a rigid set of rules but a guide for constructing shared standards together. What abstract constraints imply in context can only emerge from engagement, i.e. from application to context. What constitutes principled accessibility or comprehensibility can be sketched in the abstract with regard to the conditions that support agency and
Practical Reasoning and Transnational Justice

reason, but cannot be fully specified without deliberation between real and historically situated agents.

In this way, what O'Neill's account of the justification of moral principles offers is a universal account of practical reasoning which can guide deliberation that attempts to distinguish between reasonable and unreasonable positions. Such a position could avoid the problems of substantive, liberal positions that cannot accommodate pluralism, by offering an accessible justification for enabling constraints on deliberation and public reason. By supporting capacities and conditions for inclusive participation and legitimate agreement it can therefore delimit the scope and legitimacy of a reasonable consensus in a way that is justifiable to all humans as moral persons.

3.3. Conclusion

This analysis of public reason has attempted to demonstrate that, for Kant, public reasoning is fundamentally moral reasoning: the conditions for the possibility of dialogue and communication are also the conditions for the possibility of reason and action that Kantian morality is concerned with. Part 3 began by reviewing O'Neill's distinctive interpretation of the foundation of Kant's ethics in the "political metaphor". This critical feature of her interpretation is particularly interesting, as it offers an understanding of "political" that greatly contrasts with Rawls's understanding of a "political conception of justice". On O'Neill's interpretation a political context is designated by a plurality of finite, rational, vulnerable and mutually dependent agents who are attempting to construct principles that can guide their interactions with each other. This practical, political task can be seen as the construction of principles of justice or correspondingly as the construction of shared standards for reasoning. In this way the formulations of the Categorical Imperative point to a politically orientated starting point to ground their justification. Justification begins with the consideration that the agents of construction must at least not rule out the possibility of construction. Constraints on maxims for action arise from the consideration that they are to apply to a
plurality of agents; hence the implication that no maxim upon which others cannot act can be acceptable to a plurality.

O'Neill argued that Kant's accounts of freedom, reason, autonomy and agency together provide the framework for guiding the coordination of action. While we cannot prove human freedom, there is reason enough to assume, for the practical purpose of constructing principles, that we have a sufficient level of free agency. However, the negative use of freedom is not sufficient to ground autonomy, rather negative freedom is only freedom from things and hence only a condition of morally valuable freedom. The positive use of freedom is what constitutes moral action. Positive freedom, as self-legislation, must be understood to relate to our capacity for moral reasoning, rather than to individual selves in general. Positive freedom is the autonomy of reason: reason submits to no law other than that which it gives itself. This implies reason that constrains itself only by considering what principles or reasons could not be shared by a plurality of agents. The crucial consideration is agency: the capacity to share principles requires that one's agency be sustained. Hence respect for the agency of moral persons is the fundamental consideration for constructing shared principles of reason and justice.

O'Neill argues that several of the formulations of the Categorical Imperative are equivalent expressions of the basic capacity for morality. The Formula of the End in Itself requires that we respect the agency of others and do not instrumentalise them without their consent. O'Neill emphasises it is not consent itself that is fundamental: it is the capacity to consent, implying the underlying capacities for agency. It is this understanding that provided us with the basis for the critique of Rawls's foundational use of agreement in his account of domestic and consequently international justice. The Formula of the Universal Law, or the supreme principle of practical reason, mirrors the Formula of the End in Itself by asking not how we treat others but how we must act. It asks what principles we must reject if we aim for justification that can extend to an unrestricted audience, i.e. unconditionally. Reason directed at a restricted audience does not aim for this universal horizon and hence cannot have unconditional status.
If reasons or maxims are not in principle universally accessible, reason cannot claim authority. The Formula of the Universal Law is the formula for guiding action and for regulating reason. Maxims for thinking and acting must be adoptable, followable and intelligible to all. Anything else is heteronomous and hence lacking authority. Only autonomous reason can vindicate the authority of reason.

Autonomy, as the self-rule of reason, addresses the minimum requirement for constituting reason's authority, which (recalling the politically orientated starting point) is that discovering principles for a plurality cannot rely on a pre-established agreement. Rather the fundamental constraint on reason is whether a reason or principle is shareable by all, whether it is universalizable. Ultimately then universalizability is simply the demand that we respect the agency of those with whom we expect to interact or reason with. Autonomy means applying a recursive, reflexive evaluation to our reason and examining the shareability and hence justifiability of our reasons.

The basis of Kantian moral and public reasoning in comparison to Rawls is summarised by O'Neill as the following:

Kantian justification demands only that we reject principles on which no convergence of wills is possible. It insists only that anything that is to count as a reason must be something that can be offered and accepted, discussed and rejected even across (or disregarding) the boundaries of states, even among those who are not already fellow citizens with shared political identities. Rawls sees justice in contractualist terms; Kant views it as one very important domain of a reasoned theory of obligation. These contrasting conceptions of public reason as reciprocity and as universalizability correspond to contrasting conceptions of justice as contractualist and as obligation-based, and of the context of justice as internal to closed societies and as reaching across boundaries. (PLPR, 19)

Like for Rawls, a Kantian account of practical reason is the basis for public reasoning. However, unlike Rawls, Kant's account of practical and public reason does not suffer from the deficiencies that arise from deploying restricted and conditional conceptions of practical and public reason. Reasons that appeal fundamentally to the shared understandings or premises of particular groups cannot be accepted by "the world at large" and as such are not fit to function as public reasons. An agreement between citizens cannot be a presupposition of
reasoning about justice; hence the contract cannot form the basis for reasoning about justice. Reciprocity cannot function outside the context of the contract, and hence cannot ground public reason. Instead, Kant grounds public reason in moral obligation, in the condition of universalizability, which is to be adhered to also unilaterally. Such a condition offers a more promising structure for public reason in that it does not need to presuppose the contingencies of a particular society in order to ground a framework for public reason that is both open and appropriately constrained. Where Rawlsian public reason undermines the authority of reason through its restrictions, Kantian public reason only restricts in order to enable; it constrains only where a lack of constraint would disable communication.

In this way the justification of Kantian principles of justice is also the ground for the construction of shared standards of reasoning in an unrestricted plurality, and for further deliberation on pressing political, ethical and moral questions. Crucially, this account is only offered as a practical framework of mutual respect for further deliberation between historically contextualised agents who are otherwise free to determine the direction of their shared future. Once respect for the agency of humans as moral persons grounds discourse, we need not allow a fear of disagreement to prevent us from engaging each other in a spirit of practical and “intellectual solidarity”.
Summary and Conclusion

This thesis has aimed to demonstrate that accounts of justice that focus on whether principles are or would be accepted by particular pluralities of citizens offer insufficient justification for constructing principles of justice and shared standards for morality and reason. Probable, actual, or hypothetical forms of acceptance of principles of justice, or of particular “ideas” of practical reason, are no indication of the justice or objectivity of principles and as such are not suitable criteria for fundamental justification. Despite what Forst identified as a suggestion of the universal validity of Rawls’s particular conception of practical reason and its corresponding “ideas”, Rawls fails to justify this account in a way that is not in the end reliant on the perspective of a particular set of persons. Linking justification to what is actually accepted by particular persons in this way undercuts the unconditionality of principles of justice, making them contingent and in this sense not principles at all. Regardless of whether or not we ever could limit justification to restricted political contexts, the realities of the last several decades of global integration mean that we can certainly no longer circumscribe justice only within distinct “domestic” political spheres.

Principles are unconditional not because they receive actual universal acceptance but because they articulate what cannot be rejected without risking the very foundation of the context of justice. The moral justifiability of a principle hinges not on its neutrality toward worldviews but on whether that principle sustains or erodes the social context of humans as moral agents. The fundamental consideration is that where we assume agency in others we are obligated to respect that agency and not act in ways by which it may be damaged or eroded. Without a plurality of agents who can interact with one another, the very context of justice and justification is lost.
O’Neill’s insight, aside from asserting the necessity of constraints on reason and action over their desirability, is that principles of justice apply to actual agents. One implication is that this draws attention to the fragility and vulnerability of actual human capacities for action, and as such represents a challenge to ways of thinking that interpret justice as a formal requirement that need not consider the context of inequality, power and powerlessness that shapes all interactions. O’Neill’s account demonstrates convincingly that appeals to justice are futile where they are not translated into a real commitment to secure, foster and promote the capabilities of human agents. Inequality and justice cannot be separated: wherever certain persons lack power relative to others, injustice is almost inevitable. O’Neill’s argument rejects the idea that it is sufficient to focus on the justice of interactions in contexts of severe vulnerability, rather we must assume that where there are differentials of power there will also be abuses of power. Where it can be shown that those with whom we interact and to whom we therefore owe justification are located both within and far beyond “our” borders, these conclusions represent demanding challenges to the current status quo on questions of global justice, which currently tends toward a rejection of extensive duties of justice beyond borders. Theories of justice such as Rawls’s, that use the concept of duty in a restricted sense and focus it on the domestic society, reinforce the status quo and legitimate political hesitation and prevarication on urgently required action on poverty and inequality. It also legitimates a categorical rejection of principles of justice by those who rely on the political rhetoric of cultural relativism to suggest that freedom and equality are western principles that do not constitute obligations to other cultures and traditions. O’Neill’s account challenges this rhetoric by embedding principles of justice in an abstract and in principle accessible account of human agency and capacities for autonomous moral reasoning. As for whether this approach is a risk to the “stability” of the world order, that is a further question, though I would suggest that mere agreement, at the expense of justification and the affirmation of the equal freedom of all humans as moral persons, cannot ensure peace and stability.

O’Neill’s emphasis on actual, contextualised agents also implies that the actual experiences and self-understandings of particular agents are essential to the
Practical Reasoning and Transnational Justice

concretisation and realisation of justice. In this way her account could be brought into dialogue with thinkers from particular ethical traditions, such as the Christian ethicist David Hollenbach, who discuss the necessary realisation of universal justice in concrete contexts. O'Neill rejects accounts of justice that make rights the sole focus of obligation and that present rights as if they had to be seen in opposition to ethical questions of the common good. Her insight is that rights alone obscure the idea that we have imperfect obligations toward others, obligations that cannot be claimed yet still place demands on our action. She argues that mutually interdependent and vulnerable agents cannot thrive on perfect obligations alone; in this way her thinking could also be brought into dialogue with perspectives that question the adequacy of a restricted political account of reciprocally grounded rights in the public sphere. One of these is articulated by the Christian ethicist Duncan Forrester who argues that reciprocity is insufficient to sustain justice and lacks the motivation that inspired historical movements which brought about social change. The "measured consensualism" does not reflect the "visionary accounts of justice which have motivated most great movements for social transformation and reform".211 He points out the need to draw on richer visions of justice that include unilateral principles such as "generosity". Justice, for Forrester, cannot be achieved without a recognition of the need to reject self-interest and affirm anticipative and reconciling principles: that accounts of justice must be "robust enough to face real conflicts of interest and of understanding, and visionary enough to evoke a passionate commitment capable of calling forth self-sacrifice and challenging self-interest".212 Since reciprocity makes justice contingent upon others' action, such a principle cannot ground a real commitment to justice where "self-sacrifice" may be what is needed to open up possibilities for reconciliation and mutual recognition.

O'Neill, by articulating the necessity of (imperfect) obligations that cannot be claimed by others, offers a bridge between accounts of rights and a perspective on ethics and the common good. The resources of distinct ethical traditions, their historical contribution to thinking about justice, and the radical challenge they

---

212 Forrester, "Fairness", p. 128.
pose to the status quo, cannot be unlocked where what Hollenbach described as a “liberalism of fear” disables communication in the public sphere. O’Neill’s account supports the view that visions of the common good are compatible with a respect for the equal freedom of all persons, that these traditions sustain the contexts in which justice is concretised, and offer the motivating resources for seeking justice in conditions of solidarity with others.

Another direction of exploration to which this thesis points is a dialogue between O’Neill’s modal interpretation of Kantian principles for reason and communication and thinkers in discourse ethics, Habermas and Forst being two perspectives I have attempted to draw attention to. O’Neill emphasises that possible agreement, which implies principles for respecting agency, is more fundamental to justification than actual agreement. A question raised was to what extent this implied a right to override actual agreements, and I noted that this may not be the appropriate position from which to approach this question, rather it seems more promising to interpret her claims as a demand for action against conditions of inequality and vulnerability. O’Neill conceives of persons as moral agents first, so it is unlikely that her arguments support a disregard for the actual agreements of concrete agents.

However, the question remains of how exactly one parses the relationship between possible and actual consent or agreement, and it is in this regard that her account could benefit from dialogue with discourse ethicists. In apparent contrast to discourse ethical approaches, O’Neill does not see “inclusive” and “uncoerced” discourse as sufficient to justify agreement, though it may be sufficient to legitimate it. This would have to lead into a discussion of how to distinguish the two; similarly, for Habermas not every consensus is a true consensus. O’Neill distinguishes freedom to participate from norms of reasoning, and in this way a starting point for dialogue may be to what extent an ethics of communication focuses on the rights or the obligations of participants, and whether the rights of participation extend to considering the fundamental conditions by which participation can be sustained. This insistence on the need for conditions for
agency and participation to constrain the outcomes of discourse may help clarify what is entailed by the test of “universalization”.

Both O’Neill and discourse theories of ethics share insights on the relationship between rights and reason that make it possible to imagine a different course of argumentation: that if Rawls had not attempted to justify liberty rights by reference to what people would want or agree to, he could have connected them to their indispensability in relation to legitimate political institutions and to all public reasoning. He might then have considered how to construct an account of the ethics of communication in public reason, rather than to privatise and restrict forms of reasoning out of fear that critical intellectual engagement and deliberative democratic procedures were a threat to “our” fundamental rights.
Bibliography


Forst, Rainer. Contexts of Justice: Political Philosophy Beyond Liberalism and


---, "Kant on Reason and Religion." *The Tanner Lecture on Human Values.* Delivered at Harvard University, April 1-3, 1996.
Practical Reasoning and Transnational Justice


Regan, Ethna. Theology and The Boundary Discourse of Human Rights. Washington: 
Sen, Amartya. “Open and Closed Impartiality.” In Global Rights and Global 
Tan, Kok-Chor. “Liberal Toleration in Rawls’s Law of Peoples”. In John Rawls: 
Critical Assessments of Leading Political Philosophers. Vol. 4: Political Liberalism 
Pp. 193-211.