Philosophy Today
DOI: 10.5840/philtoday2020108355

Grounding Reasonableness in Rawls's Reading of Hobbes

KARIM BARAKAT

Abstract: I argue in this paper that Rawls is unable to offer a ground for the normativity of his freestanding politics, where his account is susceptible to a number of criticisms he raises against Hobbes. Rawls identifies three problems in Hobbes's political view: the absence of reasonableness, the lack of a social role for morality, and finally resorting to an absolute sovereign to maintain stability. I maintain that Rawls's Kantian account circumvents these problems. However, I argue that his move to a freestanding politics that disposes of the Kantian moral basis is unable to justify normative commitments and ultimately resorts to contingent justifications resulting from uncritically accepting norms institutions inculcate.

Key words: Rawls, Hobbes, social contract, reasonableness, freestanding politics

The resurgence of social contract theory following the publication of *A Theory of Justice* has led to focusing on hypothetical contracts as a means to justify political stability. Rawls develops a procedural political account that relies on a hypothetical agreement in order to arrive at the basic structure of institutions. To account for political justification, Rawls begins by rejecting a Hobbesian view that inevitably leads to an absolute sovereign. He takes Hobbes's view on contractarianism to be inadequate in terms of its failure to account for stability for the right reasons, the absence of reasonable incentives, and the lack of a social role of morality. Instead, Rawls advances a version of the contract that he labels initially as Kantian. This turn to Kant affords him the notion of moral persons with powers capable of acquiring principles of justice due to a learning process. Rawls, however, soon realizes that this view entails exclusionary commitments that render his account inconsistent. Consequently, he chooses to abandon this approach in favor of a freestanding political view.

ISSN 0031-8256

This paper will focus on the extent to which Rawls is able to ground his political account following the problems he identifies with Hobbes. I begin with the reading Rawls offers of Hobbes and focus on these objections as motivating his general account, examining first Rawls's *Lectures on the History of Political Philosophy*, in order to identify three shortcomings that he finds in the Hobbesian characterization of practical reason and the state of nature. I then take up Rawls's Kantian move to show how he attempts to resolve these problems while shedding light on the dependence of this view on moral conceptions that are not reducible to politics. I argue that Rawls's Kantian view, however, is susceptible to the problem of feasibility Sen puts forward. In *The Idea of Justice*, Sen casts doubt on the possibility of Rawls's goal of arriving at a coherent set of principles of justice. Moreover, Sen argues that Rawls's view does not address the question of how we ought to realize socially this ideal.[1] I, then, shed light on the amendments Rawls introduces in order to develop a freestanding political account that manages to partially respond to doubts concerning feasibility. I argue, however, that Rawls's view fails in three respects. First, Rawls's inability to offer a standard for determining the reasonableness of citizens and doctrines leads to one of two conclusions. Either his view is not to be regarded as one claiming universal validity, or it will necessarily be exclusionary to other liberal views. Both conclusions are not ones Rawls would be willing to accept. Second, the view of moral psychology to which he subscribes cannot be understood as belonging to an independent sphere but must be justified by appealing to a comprehensive doctrine. Finally, I argue that Rawls overlooks the problematic role institutions play, by his own lights, in inculcating liberal ideals in individuals and thus violates his commitment to the "freedom" of citizens in the original position. Ultimately, I argue that developing a freestanding politics fails to address the problems Rawls locates in Hobbes's view.

Rawls's Reading of Hobbes

In *A Theory of Justice*, Rawls identifies his account as contractarian, one that finds affinities with the views of Locke, Rousseau, and Kant, while distancing itself from Hobbes. Rawls states,

> I shall regard Locke's *Second Treatise of Government*, Rousseau's *The Social Contract*, and Kant's ethical works beginning with *The Foundations of the Metaphysics of Morals* as definitive of the contract tradition. For all of its greatness, Hobbes's *Leviathan* raises special problems.[2]

These "special problems" Rawls identifies with *Leviathan* lead him towards articulating a liberal democratic approach that promises to avoid introducing an absolute sovereign as a necessary condition for the success of a political view. Consequently, the soundness of Rawls's position can be partly determined by how well he manages to resolve these problems he identifies in Hobbes.

We find Rawls's criticisms of Hobbes most clearly stated in his *Lectures on the History of Political Philosophy*. Rawls begins by arguing that Hobbes's state of nature ought to be interpreted as hypothetical, where rational individuals move away from the condition of war that would only result in the absence of an absolute sovereign. Rawls takes Hobbes to argue that if we examine the way human beings act in society, we would be able to understand how they would act in the state of nature. While referring to Hobbes's argument from experience in the state of nature, Rawls contends that basic self-interested desires not only remain active but are also not susceptible to transformation in the presence of social institutions. Therefore, the basic traits leading to conflict we find in the state of nature are essential characteristics of human beings.[3] This leads us to the first main problem Rawls locates in Hobbes's view. His contract does not give an account of how politics, or morality for that matter, plays a role in shaping individuals.

Having specified basic unchanging human desires as the basis of Hobbes's argument, Rawls raises the question of motivation and practical reason in Hobbes. Hobbes's laws of nature involve demands of reciprocity that guarantee social cooperation.[4] However, Hobbes justifies those principles by appealing entirely to individual self-preservation and thus self-interest. Yet, for Rawls, this omits a capacity for reasonableness that is needed to arrive at social cooperation without coercion. Rawls begins by distinguishing the rational from the reasonable, a distinction he argues is not present in Hobbes. Whereas rationality is defined in terms of satisfying one's own interests, reasonableness involves "fair terms of cooperation."[5] Though fair cooperation develops out of working for the rational advantage of one's self, and in this sense always presupposes rationality, reasonableness requires an additional component, that of justice arising out of reciprocal relations. This leads Rawls to introduce a major argument against Hobbes, namely, lacking grounds for moral obligations and political rights.[6] For Rawls, Hobbes fails to offer grounds for reasonableness independently of self-interest. The relevant question Rawls attempts to tackle, then, is how to provide a constructivist account that would simultaneously be able to offer reasonable grounds for obligation. This constitutes the second problem Rawls locates in Hobbes's position.

The problem of the grounds of reasonableness paves the way for an additional problem Rawls raises against Hobbes's view, that of developing stability. As is explicit in Hobbes's account, the sovereign plays the role of "stabiliz[ing]" the contract. What guarantees that we eventually arrive at stability in Hobbes's situation, where covenants can be relied upon and agreements hold, is an absolute ruler. In the absence of a sovereign, the contract will not be observed, and we return to the state of nature.[7] The sovereign, therefore, changes what Rawls refers to as the "background conditions" to which individuals refer in their rational decisions. Rather than changing the psychological traits that lead to conflict, the presence of a sovereign ensures that it becomes *rational* to obey the laws of nature.

Whereas Hobbes considers the condition of absolute sovereignty to be necessary for moving towards a stable community, Rawls rightly argues that this claim can be disputed on mere historical grounds. As Rawls maintains, "we *know*, I think, that Hobbes's substantive theory can't be, *in general*, correct; since *constitutional democratic* institutions that violate his conditions for the Sovereign *have actually* existed and have not been noticeably less stable and orderly regimes than the kind of absolutism that Hobbes favored."[8] Accordingly, stability under the constitutional system Rawls espouses must exist alongside a democratic procedure rather than being guaranteed by absolute power. Rawls maintains that reasonable self-restraint and fairness are essential to social cooperation, which involves reciprocity and willingness to abide by one's obligations on the condition that others do so too. For Rawls, "[t]hese *reasonable desires* . . . have no part in [Hobbes's] account of human psychology, at least insofar as political questions are concerned."[9] Even if such desires exist in Hobbes's scheme, they are simply ineffective and unreliable. Accordingly, Rawls presents his project as attempting to "recast [the Social Contract view] so as to provide . . . a framework within which the content of the notions essential to social cooperation—reasonable self-restraint and fairness—can be defined or outlined."[10] Though Rawls finds that these conditions are partially satisfied in Locke's view, it is ultimately in Kant and the concept of a moral person that his account begins to address these problems before finally developing an independent political view.

From Kantian Constructivism to Freestanding Politics

Rawls contends in the preface to *Theory* that the theory he develops is "highly Kantian in nature."[11] He later emphasizes the Kantian undertones of his view in several essays following the publication of *Theory*. In "A Kantian Conception of Equality" and "Kantian Constructivism in Moral Theory," Rawls outlines the procedural starting points of his view, while shedding light on the relation between equality and Kantian freedom. He specifies two main ideals with which we procedurally begin, that of the well-ordered society, along with the second ideal of free and equal moral persons. This is necessary because for him political theory is primarily concerned with constructing the basis for resolving a *moral* question and thereby requires providing an account of moral persons possessing specific formal qualities. In addition, the ideal of a well-ordered society is one whose members accept and know the public conception of justice, whose social institutions express the principles of justice, and finally where this conception is founded on methods of inquiry that have been agreed upon. Beginning from this liberal democratic ideal, stability will ensue due to the perpetuating the principles of justice through public expression. This in turn allows for reaffirming the commitment of citizens to this specific understanding of justice.[12]

Yet, to see how this view responds to the problems Rawls locates in Hobbes, it is essential to develop further his Kantian account. Central to his thesis is the application of Kantian ethics to the social sphere.[13] Rawls conceives the fundamental political question as one arising out of a philosophical incompatibility between freedom and equality. The question that follows is how to develop a political view that maintains both freedom and equality which Rawls contends are necessary for a well-ordered society. Rawls, therefore, introduces a hypothetical original position, in which conflict arises from different conceptions of the good. Resolving the problem of difference requires an agreement on the basic principles of justice. Hypothetical agreement, thus, serves as the basis for *constructing* the principles of justice under conditions that render personal preferences opaque. Kantian constructivism, then, emphasizes the role of agreement, and "holds that moral objectivity is to be understood in terms of a suitably constructed social point of view that all can accept."[14] The practical nature of the political problem requires that we introduce a procedure that allows us to arrive at the most reasonable understanding of justice that all reasonable groups accept. Whereas this position is not grounded metaphysically, Rawls chooses a moral ground for developing his politics. Under his Kantian view, the formal conceptions of moral persons as free and equal hypothetically set aside their different conceptions of the good in order to shed light on shared principles.

Having started with basic commonsensical conceptions of ideal social relations (a well-ordered society) and moral persons, we are able to identify features that delineate the relation between individuals and society under conditions of freedom and equality. Moreover, moral persons are to choose *ideal* principles of justice to construct their well-ordered society independently of arbitrary distributions of wealth and power, or under a veil of ignorance. Specifying an understanding of a moral person becomes essential for Rawls's project. In a Kantian vein, such persons must first be rationally autonomous agents taken in abstract from their daily civil roles.[15] What makes these persons moral is, first, their "capacity to form, to revise, and rationally pursue a conception of the good," or what Rawls refers to as rationality, and second, recognizing that others are moved by similar motives of realizing their respective conceptions of the good.[16] Consequently, Rawls embeds the notion of reasonableness he had argued was missing in Hobbes in the concept of a moral person. Accordingly, moral personality also requires that persons have a "sense of justice," and are able to act out of the principles of justice, or reasonableness. In fact, Rawls maintains that acting *out of* a sense of justice is a necessary requirement to arrive at stability.[17]

Reasonableness, moreover, sets limits on rationality since it determines the boundaries for pursuing one's conception of the good in relation to other members of the community. This limiting function of reasonableness emphasizes a key feature of Rawls's Kantianism, for "[t]he priority of the right over the good

is characteristic of Kantian constructivism."[18] The Kantian view Rawls endorses, therefore, departs from his picture of Hobbes's position by maintaining that reasonableness is justified on its own grounds, through appealing to a moral capacity that is not reducible to calculative reason or self-interest.[19]

Understanding moral persons engaged in a political agreement as reasonable provides the grounds for addressing two further problems Rawls finds in Hobbes, namely, moral development and political stability, which he argues are inseparable. To arrive at stability, Rawls invokes the concept of "moral learning" which guarantees that persons "develop a desire to act" following the principles of justice. The idea of moral learning stresses the social role of political institutions. Furthermore, it depends on a psychological understanding of human nature in order to determine whether we can expect citizens to learn to act out of a set of principles. Ultimately, this rests on publicly recognizing the basic principles of justice, which for Rawls leads citizens to act out of these principles.[20] Consequently, given the specific conception of moral persons, Rawls's view accounts for the reasonableness of individuals, the social role of political institutions, and finally ensuring that society will be stable as a result of this social role.

Nevertheless, grounding politics on moral presuppositions raises questions concerning the feasibility of Rawls's view. In *The Idea of Justice*, Amartya Sen contends Rawls offers no guarantee that an agreement will be reached upon a unique set of principles of justice. Sen distinguishes between what he takes to be Rawls's contractarian "transcendental institutionalism" on the one hand, and his "realization-focused comparison" on the other.[21] The former is the result of defining a perfect sense of justice that is to be realized institutionally, one that does not begin with the question of the social realization of this ideal.

According to Sen, this raises concerns regarding the feasibility of transcendental institutionalist views, as it is unclear whether an agreement on a single set of principles of justice would in fact ensue. A political theory must, therefore, be able to incorporate several, possibly incompatible, conceptions of justice. Furthermore, Sen maintains that the transcendental solution is redundant. Given its ideality, it does not render superfluous the need for selecting a feasible solution among different applicable alternatives.[22] The problem of the social realization of a conception of justice hinges on a distinction between agreement and actual realization. Sen maintains that the attainment of justice must be the product of a "gradual formation of behavior patterns" among citizens, which do not appear immediately upon agreement.[23]

Sen takes his critique to apply to Rawls's work generally. However, he refers primarily to *Theory* and appears to overlook for the most part the freestanding position Rawls adopts in *Political Liberalism*. A freestanding political conception, according to Rawls, is not derived from a comprehensive doctrine, and though it is moral, it "must contain its own intrinsic normative and moral ideal."[24] The view

that Rawls introduces in *Theory* is significantly modified in his later political turn. In *Political Liberalism*, Rawls maintains the notions of rationality and reasonableness, while severing the connection with a Kantian conception of moral persons.[25]

By appealing to a freestanding political view, Rawls is able to partially respond to the objections Sen raises once the moral ground has been discarded. In his "Political Liberalism: Reply to Habermas," Rawls argues that unlike Habermas who is concerned with identifying the basic conditions of communication, Rawls introduces the minimal basic requirements for a democratic procedure. Rawls maintains that the view he advances in *Political Liberalism*, like Habermas, does not presuppose a Platonic or Kantian reason, but rather a freestanding notion. Rawls adds "[n]o sensible view can possibly get by without the reasonable and rational as I use them."[26] Following the reading I have suggested, Rawls would argue that these two features of citizens constitute necessary conditions that make it possible to have a liberal democracy. Rawls locates these conditions in rational and reasonable citizens.[27] Under this reading, justice as fairness becomes at least partially feasible, given that we only need to accept that liberal democracy is possible for the view to obtain. Once this is granted, Rawls's view unfolds as an analysis of such necessary conditions. It is still questionable, however, whether we need some minimal agreement to begin with for a democratic procedure to be possible.

Nonetheless, by setting aside the notion of moral persons, Rawls's account appears to fail to respond to at least one problem he identifies with Hobbes's view, that of realizing the conception of justice socially. In fact, though Rawls invokes a "theory of moral learning" in *Theory*, his later *Political Liberalism* does not make use of such a notion at all.[28] *The Law of Peoples*, alternatively, appeals to moral learning cursorily as a needed process for the realization of an agreement on international principles of justice, or among nonliberal but somewhat reasonable societies.[29] Liberal societies, alternatively already "must have political and social institutions that effectively lead its citizens to acquire the appropriate sense of justice as they grow up and take part in society."[30] In fact, one main reason for not introducing moral learning in *Political Liberalism* can be explained by Rawls's claim that liberal democracies already have an overlapping consensus over the values that must be learned. This claim of an already existing overlapping consensus in society minimizes the role of the process of moral learning in the political sphere. This leaves open the question of how Rawls can put forward a view of the social role of his political conception of justice if he disposes of such a psychological process, a problem I will return to in the next section.

But there is another reason for not appealing to moral learning in *Political Liberalism*. In *Theory*, Rawls maintains that moral learning "depends upon an account of the nature of morality and its various forms."[31] Rawls's earlier view on moral learning, thus, can only be part of a comprehensive doctrine, which he comes to realize is inadmissible into a strictly political account.[32] Rawls still provides a

political view on moral psychology, arguing that citizens are amenable to the view of liberalism he suggests, but the political moral psychology he puts forth requires a modified *political* view on moral learning to complement it.

This clarification of Rawls's position provides further means to respond to the problem of feasibility Sen raises. Sen had expressed his skepticism with regards to arriving at a consistent set of principles of justice under the veil of ignorance. Sen later makes a reference to *The Law of Peoples*, where Rawls appears to mitigate the requirement of a basic agreement between citizens in terms of a "family of reasonable political conceptions," as opposed to a single political conception.[33] He rightly points out the tension between this view endorsed in *The Law of Peoples* and the one already presented in *Theory*. The inconsistency between the two views Rawls advances, however, should be explained by the shift away from Kantianism we find in Rawls's later works. In *Political Liberalism*, Rawls explicitly distances the view he presents from the one he had introduced in *Theory*, where he had explicitly endorsed a comprehensive doctrine. Rawls's freestanding politics aims at uncovering an already-existing agreement with regard to the basic structure of institutions. Rawls is, thus, committed to a minimal requirement for agreement in *Political Liberalism*, one that he maintains follows immediately from reasonableness and rationality taken as moral regulative ideas.[34] This minimal requirement arises from his commitment to the reasonable fact of pluralism along with irresolvable causes of difference, or what Rawls refers to as the "burdens of judgment."[35] The fact that agreement over the basic structure of social institutions is already present renders Rawls's view more flexible. Challenging this claim requires showing that even such a minimal agreement is not available or possible, an argument that Sen does not provide.

Under this reading, the question of the redundancy of "transcendental institutionalism" introduced by Sen will depend on the relation between institutions and society Rawls puts forward for realizing his conception of justice. It is noteworthy, however, that if we accept Sen's argument that Rawls's position is not feasible due to its ideality, then at worst, Rawls appears to offer the normative grounds for selecting between different options based on a political ideal.[36] Rawls, however, maintains that this ideal can be realized at least to some degree. Nonetheless, though I have argued that Rawls's account is more feasible than Sen suggests, it is still unclear how the social role of such a political account can be achieved in order to realize the conception of justice. Moreover, in the absence of moral presuppositions, the normative content of his view will need a different kind of justification. Responding to these issues requires exploring Rawls's view of publicity and stability through shedding light on the idea of an overlapping consensus as I will proceed to argue in the next section.

The Inadequacy of the Political

The previous section introduced questions concerning the feasibility of Rawls's Kantian account while emphasizing the modifications following the shift from a Kantian conception of justice to a freestanding one. Rawls's attempt to render his view feasible involves specifying social conditions that are required for the realization of justice as fairness. I have argued that responding to the problems Sen raises requires shedding light on the account of the process required to instill the principles of justice in citizens. Though this requirement is present in *Theory*, Rawls does not make a reference to such a process in *Political Liberalism*. In order to provide a more complete response to the problem Sen raises, this section focuses on the other two shortcomings Rawls identifies in Hobbes's account, namely stability and the social role of institutions. Rawls attempts to account for such phenomena by presenting a freestanding political view that renders his account consistent with various reasonable comprehensive doctrines. This requires treating politics as entirely separate from non-political grounds including material ones having to do with how the social role is manifested. Yet, whether this freestanding view is adequate rests on being able to introduce a sufficient justification for the stability of the political scheme.

Rawls introduces different views on how stability is achieved between *Theory* and *Political Liberalism*. In the former, stability appears to result from the process of moral learning Rawls specifies, which is justified by the moral conception of individuals. Moral learning occurs due to the satisfaction of the publicity requirement Rawls introduces. In contrast, in *Political Liberalism*, Rawls retains the public function of justification but appeals primarily to the political idea of an overlapping consensus to account for the acquisition of liberal democratic values by citizens. In relation to Rawls's reading of Hobbes, these arguments allow him to provide a social role for his political conception and maintain stability for the right reasons.

Clarifying how stability is attained and what social role political institutions can perform requires first fleshing out Rawls's separation of politics from metaphysics, especially when it comes to the capacities of citizens. Rawls argues against any metaphysical dependencies by referring to his freestanding political account of moral psychology. For Rawls, political persons are ready to bind themselves to specific principles due to their desire to be part of a well-ordered society. As Rawls notes, this account is political, neither philosophical nor psychological, as it is "drawn from the political conception of justice as fairness" and "express[es] a certain political conception of the person and an ideal of citizenship."[37] This last raises the question of how we arrive at this notion of citizens. One way to justify this specific ideal of citizens along with the moral psychology Rawls assigns to them is through reversing the procedure, beginning from a well-ordered society in

order to derive a concept of an ideal citizen. Accordingly, from a political point of view, these are the conditions necessary for a well-ordered society to be realized.

But whether Rawls is justified in attributing to these commitments a universal status remains unclear. In the absence of a metaphysical ground and given Rawls's explicit claim that in his theory political concepts are those of a liberal democratic state, one can argue that Rawls's conclusions end up being limited in scope, and do not acquire universal validity.[38] This is further emphasized with his political turn. In fact, in *The Law of Peoples*, Rawls considers these ideals to be consistent with his specific conception of reasonableness of a liberal democratic society. The degree to which a community (and this applies to individuals as well) is reasonable determines the degree to which those ideals apply.[39] Again, Rawls's argument requires that one first begin with the liberal democratic model. But as O'Neill argues, Rawls comes to assume the legitimacy of liberal democracy and does not argue for it.[40] Rawls, thus, seems to dispose of universality in exchange for limiting his view within the confines of a political tradition.

Robert Taylor raises another problem for the justice as fairness account, arguing that the limits Rawls imposes on his view entails a "pessimism about the width of reflective equilibrium," given that our conclusions are determined by the traditions from which we begin.[41] If the procedure we prescribe will heavily depend on adopting a liberal democratic view, it is not clear how Rawls can still maintain that wide reflective equilibrium, or whether arriving at a coherent view as a result of consulting the conceptions of others, is possible. As Taylor argues, "[s]trong path dependence in reflection turns all reflective equilibria into narrow ones" (italics in original removed).[42]

It is not entirely clear, however, why this constitutes a problem for Rawls's position. Rawls can simply reply by arguing that reflective equilibrium, especially wide reflective equilibrium, will only be properly effective when conditions of publicity are met. Moreover, these conditions are in fact satisfied in a liberal tradition. In the absence of publicity, one simply cannot negotiate one's own views with those of others, which renders reflective equilibrium only possible in a narrow sense. In a liberal society, in contrast, the fact of reasonable pluralism guarantees that reflective equilibrium will not be merely narrow. Despite the existing overlapping consensus on basic matters, several reasonable comprehensive doctrines will still function as sources for a plurality of possible worldviews.

Arguing that liberalism is a precondition for the success of the procedures of justice as fairness provides Rawls with another way to respond to the question of universal validity. As I have suggested, Rawls resorts to an actual overlapping consensus instead of moral learning in *Political Liberalism* since the problem of *actually* constituting subjects who are acting out of a sense of justice becomes less acute if we begin with individuals with reasonable comprehensive doctrines in a liberal democratic society. By making use of the overlapping consensus existing

between the different publicly available reasonable comprehensive doctrines, we can begin with the already agreed upon conception of a well-ordered society and accordingly procedurally construct the principles of justice. The notion of a well-ordered society invoked will be based on the overlap between different comprehensive doctrines. Justifying the procedure, under this argument, is not simply based on following an ideal liberal tradition, but on the actual existence of a consensus from which we can begin. There is certainly some incompleteness in Rawls's view with respect to arguing for the universal validity of liberalism, but his view does not amount to a rejection of such universal validity.[43] In fact, one could argue that the superiority of Rawls's view over others would be established if we accept his claim that the fact of pluralism requires that we find basic points of agreement. While emphasizing the procedural nature of the political in attempting to resolve the practical problem of difference, Rawls could maintain his position offers a way to do that based on agreement on basic principles of justice. But Rawls never develops an explicit argument for the advantage of his position based on agreement over that of others, which renders his position incomplete. However, this does not necessarily indicate that his view is dependent on nonpolitical notions or historically contingent traditions.

This reading, however, has one serious implication; liberal views that do not subscribe to Rawls's account of a well-ordered society are rendered automatically unreasonable. To see how this follows, it is first imperative to look at one problem raised by Habermas, who argues that Rawls equivocates on the conditions that determine reasonableness. Whereas reasonableness is presupposed for any agreement, the relevant criteria for determining reasonable doctrines can either be through appealing to an already existing overlapping consensus, or it can be justified by the comprehensive doctrines of individuals. Clearly, Rawls would want to avoid committing to the latter claim given that it cannot provide an independent standard for excluding unreasonable comprehensive doctrines, and thus reasonableness will have to be justified on the basis of an existing overlapping consensus. The problem, however, is that the considered overlap must be only between individuals with reasonable comprehensive doctrines. The overlap itself, therefore, cannot be the basis to determine reasonableness, as it already presupposes a conception of the reasonable that determines whose goals and aims count to begin with.[44]

Of course, Rawls might appeal to the reading I suggested, by arguing that reasonableness can be derived from the ideal of a well-ordered society presented and not from the existing overlap. Moreover, Rawls maintains that his notion of reasonableness is a straightforward one that no theory can do without. Two problems follow from this. As I have suggested earlier, Rawls can insist on the universal validity of his view if he maintains that the idea of a well-ordered society arises out of the overlapping consensus and not as an ideal of liberalism. If that

is the case, then reasonableness cannot be based on the notion of a well-ordered society given that an overlapping consensus already presupposes a criterion for reasonableness. Instead, reasonableness must already be determined to justify a specific notion of an ideal society. Second, even if we derive the idea of the reasonable from the notion of a well-ordered society, this notion of reasonableness becomes compelling only if we accept Rawls's view of an ideal society. If we link the idea of reasonableness with the notion of a well-ordered society analytically, it follows that anyone not committed to the same idea of a well-ordered society (which cannot be derived from an overlapping consensus), or to the view of stability as Rawls presents it, will have a different conception of reasonableness. Different conceptions of reasonableness will, thus, be excluded from Rawls's reasonable comprehensive doctrines. Rawls introduces his political liberalism as a view that is able to incorporate under it many different reasonable views as candidates for determining principles of justice. But, at the meta-level, Rawls will either recognize his view as involving a local argument or will continue to maintain universality while implicitly excluding any view that is not committed to the same conception of an ideal society.

The problem that arises with the justification of reasonableness leads to a second but similar problem with Rawls's moral psychology. As we have seen, to account for stability in a well-ordered society, Rawls relies on overlapping consensus along with the moral psychology he develops.[45] Yet, the question then is which moral psychology should we accept? In *Political Liberalism*, Rawls argues that his specific political moral psychology is conducive for constructing the kind of society he identifies as well-ordered. This also involves rejecting accounts that do not quite fit with his preconceptions. For Rawls, this involves explicitly rejecting a Humean moral psychology.[46] But this has a twofold implication. First, unless Rawls can provide *political* reasons for why a Humean account is unacceptable, Rawls's view is no longer based on freestanding politics but presupposes the validity of certain comprehensive doctrines and not others. Second, given that we cannot deem any moral psychology following a Humean line of argument to be unreasonable, Rawls's moral psychology cannot itself be the product of an overlapping consensus between reasonable groups.

The discussion of the conditions of stability leads us back to the problem pointed out by Sen who argues that Rawls's view is lacking when it comes to accounting for the social realization of the ideal of justice. This ties directly to the notions of an overlapping consensus along with the condition of publicity. In *Theory*, Rawls had argued that publicity follows immediately from recognizing moral principles in terms of rational choices that govern one's conduct in an "ethical community" or "kingdom of ends."[47] This provides him with an objective criterion for determining reasonableness, albeit one that follows from a Kantian comprehensive doctrine. But in *Political Liberalism*, Rawls specifies publicity as

strictly political, where the goal of Political Constructivism is identified in terms of offering an account of the "public basis of justification . . . given the fact of reasonable pluralism."[48] Rawls argues that knowledge of the commitment of citizens to the principles of justice, general beliefs about human nature and the role of institutions, along with a complete justification for the principles of justice should be public.[49] Significantly, full publicity constitutes a precondition for freedom as only upon being familiar with justifications can citizens in fact act *out of* principles of justice. Full publicity allows for adjudicating between conflicting conceptions of freedom and equality, and since principles of justice provide the basis for institutions that shape individuals and thus have long-term effects, publicity allows for scrutiny of legal practices when necessary. Consequently, "free and equal persons are in a position to know and accept the background social influences that shape their conception of themselves as persons, as well as their character and conception of their good."[50]

The notion of publicity is, therefore, linked to the kind of justification that is given in support of one's commitments. Rather than relying on comprehensive doctrines, publicity requires that political reasons are provided for views that are endorsed publicly. This condition allows Rawls to assign a significant social role to his political account, one that he considers to be lacking in Hobbes. Rawls had argued that Hobbes assigns to individuals in the state of nature essential traits that make war inevitable. Under Rawls's reading, Hobbes minimizes the role of institutions in shaping individuals since these traits, along with the threat of war, are simply outweighed by fear of the sovereign with the transition to society. By relying on the condition of publicity, however, Rawls can account for the social role of political institutions. First, justifying the rights of citizens institutionally will directly appeal to the political publicly available conceptions of citizens. Second, through recognizing their rights as the result of their moral powers, citizens will be "educated" to understand themselves as autonomous individuals and develop the desire to be that kind of person. Rawls considers that, in a liberal democratic society, such desires would have already been at least partially nurtured, which justifies the existence of an overlapping consensus.[51] Importantly, then, citizens can recognize themselves as autonomous legislators only when this role of institutions has been properly fulfilled.[52]

A well-ordered society would, therefore, be founded on principles of justice which everyone accepts, and which have been constituted out of the reasonable beliefs of free and equal citizens. Such a society would also be stable, for moral persons that act out of a sense of justice, and where social institutions perpetuate such principles of justice, would "generate their own support" as long as these principles are "publicly realized."[53]

The influence allotted to institutions in shaping the overlapping consensus leads to a third problem in Rawls's position. Rawls recognizes that the influence

of political institutions is "pervasive" and "long-term."[54] These institutions have the effect of instilling the right kind of motivations in citizens in order to act out of a sense of justice. Though, in a liberal democratic society, Rawls claims that such motivations would have already been nurtured, he still provides a historical account of how such an overlapping consensus with regards to the principles of justice may arise. Taking the example of the development of consensus in the United States, Rawls shows how constitutional consensus eventually develops into a consensus over the actual principles of justice.[55] A precondition for this overlapping consensus, thus, is the role institutions play in order to bring into being citizens who want to take part in a liberal democratic ideal.

This points to another equivocation in Rawls's view concerning the legitimacy of the principles of justice he provides. Rawls maintains that the overlap is present among different comprehensive doctrines, and the role institutions play concerns changing the kind of justification given from one relying on comprehensive doctrines to political. But he simultaneously recognizes that the content of the overlap should be instilled by institutions. This does not simply apply to the kind of justification employed, that is, political as opposed to comprehensive, but also to the principles themselves. For instance, Rawls shows how "constitutional principles such as liberty of conscience were adopted only with great reluctance as part of a modus vivendi," before developing into an overlapping consensus.[56] Yet, the transition from constitutional consensus to an actual overlapping consensus occurs as a result of propagating these principles through political institutions. If the political view Rawls espouses is primarily justified by an actual overlapping consensus, this consensus is primarily due to the role institutions play in shaping individuals to accept such principles and ideals.

Rawls could again argue that the justification for these ideals follows from the original position given free and equal citizens and a specific conception of a well-ordered society. But as I have argued, either the notion of a well-ordered society is completely ungrounded and exclusionary of other views of an ideal society, or it is based on the actual overlapping consensus, since it is presupposed for the original position. If it is the latter, then the justification for the idea of an ideal society, ideal citizens, and the principles follows from an overlapping consensus that has been instilled institutionally. This problem is rendered more significant given that Rawls maintains that the condition of publicity is supposed to render citizens wary of the effects of political institutions. But this undermines Rawls's entire account as the specific normative ideals that publicity is designed to protect have been instilled by institutions to begin with. Rawls's argument for the specific procedure he specifies is, thus, imbued with circularity which threatens not only the normativity of his view, but also the very basic notion of negative freedom to which he is committed.

Conclusions

We can now return to the three problems Rawls identifies in his reading of Hobbes. First, I have argued that though Rawls does introduce reasonableness as independent of instrumental reason, the criteria for distinguishing between the reasonable and the unreasonable do not seem to be adequately specified. Whereas Kantian constructivism can refer to an independent nonpolitical moral basis in order to determine what is reasonable and what is not, this advantage is lost by attempting to offer a freestanding political justification for reasonableness. Consequently, Rawls will have to appeal to some nonpolitical grounding in order to justify his conception of reasonableness. Second, Rawls specifies two main requirements for stability, an overlapping consensus and a political moral psychology. As I have argued, Rawls's moral psychology cannot be understood as being strictly political and follows from commitments that cannot be justified by simply appealing to a well-ordered society. Attempting to limit his account to the political sphere results in circularity. Alternatively, Rawls would have to exclude from the outset views that do not share with him the same conception of an ideal society and rely on a different moral psychology. Finally, the notion of an overlapping consensus leads to the third problem, namely the social role of the political that Rawls finds lacking in Hobbes. Rawls, however, appears to be caught in a contradiction as he argues for grounding normativity in an original position where citizens are free, while simultaneously recognizing that the ideals to which reasonable citizens aspire, and which constitute an overlapping consensus, must be instilled by institutions. As I have maintained, one way to evaluate Rawls's position is to determine the extent to which he manages to account for the three problems he argues are present in Hobbes. I have argued that Rawls's attempt to respond to these problems ultimately appeals to normative demands that remain unjustified.

American University of Beirut

Notes

1. Sen, *The Idea of Justice*, 6–7.
2. Rawls, *A Theory of Justice*, 11n.
3 Rawls, *Lectures on the History of Political Philosophy*, 41–42.
4. Rawls, *Lectures on the History of Political Philosophy*, 61.
5. Ibid., 54.
6. Ibid., 66.
7. Ibid., 77–78.
8. Ibid., 85.
9. Ibid., 88.
10. Ibid., 88.
11. Rawls, *A Theory of Justice*, viii.

12. Rawls, "A Kantian Conception of Equality," 255–56.
13. To present this view as Kantian raises the problem of compatibility between justice as fairness and Kant's politics. But Rawls recognizes this departure, and he is quick to note that his view is not identical to Kant's. To label it as Kantian "expresses analogy and not identity," being a "doctrine [that] sufficiently resembles Kant's in enough fundamental respects so that it is far closer to his view than to the other traditional moral conceptions." In fact, justice as fairness uses certain Kantian notions to develop a political account, while disposing of the overarching framework we find in Kant's politics. The extent to which Rawls's view can be considered Kantian has been subject to criticism by a number of commentators such as Onora O'Neill and Larry Krasnoff. Alternatively, in "Rawls and Kantian Constructivism," Kaufman highlights several overlaps between the view that Rawls introduces and that of Kant. I will be less concerned with the actual overlap between the two and more with introducing Rawls's position coherently. Rawls, "Kantian Constructivism in Moral Theory," 305; O'Neill, "Constructivism in Rawls and Kant," 351; Kaufman, "Rawls and Kantian Constructivism," 229–35; Krasnoff, "How Kantian is Constructivism?"
14. Rawls, "Kantian Constructivism in Moral Theory," 307.
15. Rawls, "Kantian Constructivism in Moral Theory," 307–08.
16. Rawls, "Kantian Constructivism in Moral Theory," 312.
17. This partly addresses the problem Rawls locates in Hobbes, but fully accounting for stability will require discussing Rawls's notion of an overlapping consensus, which I will return to in the next section.
18. Rawls, "Kantian Constructivism in Moral Theory," 319. This view is also expressed in *Political Liberalism*, 52.
19. Rawls's account of the reasonable varies throughout his work. As Onora O'Neill argues, there are at last two different conceptions of reasonableness that we find articulated in *Theory* and *Political Liberalism*. O'Neill also shows that Rawls's understanding of "Kantian Constructivism" also varies between *Theory*, "Kantian Constructivism in Moral Theory," and *Political Liberalism*. O'Neill, "Constructivism in Kant and Rawls."
20. Rawls, *A Theory of Justice*, 138, 145, 177.
21. Sen also offers another argument, namely that Rawls limits his political view to the question of just institutions and disregards the role of non-institutional factors in social interactions. This is certainly an important objection, but it is outside the scope of this paper. Sen, *The Idea of Justice*, 6–7.
22. Sen, *The Idea of Justice*, 9, 98–100.
23. Sen, *The Idea of Justice*, 70.
24. Intrinsically entailing a moral ideal means this political conception is moral, but in the absence of a comprehensive doctrine, rather than basing politics on morality, this suggests that the moral conception is itself justified politically. As I argue later in this paper, the moral conception must follow from his *political* notions of a well-ordered society or from overlapping consensus. Rawls, *Political Liberalism*, xliv.
25. Rawls, *Political Liberalism*, xlv.
26. Rawls, *Political Liberalism*, 380.
27. Rawls, *Political Liberalism*, 381–82.
28. Rawls, *A Theory of Justice*, 430.

29. Rawls, *The Law of Peoples*, 66, 71.
30. Rawls, *The Law of Peoples*, 15.
31. Rawls, *Theory*, 430.
32. Of course, Rawls can develop a view of moral learning that is merely political, which is something he makes note of in *The Law of Peoples*, but this is somewhat superfluous in *Political Liberalism* as I will argue in the next section. Rawls, *The Law of Peoples*, 15.
33. Sen, *The Idea of Justice*, 11.
34. Rawls considers that the view he introduces in *Theory* requires that all citizens have the same Kantian comprehensive doctrine. In *Political Liberalism*, he distances himself from this claim. This suggests that the agreement under a freestanding political view must be far more minimal than the one he had argued for in *Theory*. Rawls, *Political Liberalism*, xlii.
35. Rawls, *Political Liberalism*, 54–55.
36. The normative basis is indeed a problem, one that I will return to shortly. Sen, however, does not dispute the normative nature of Rawls's view. Instead, he focuses on ideality as problematic.
37. Rawls, *Political Liberalism*, 86–87.
38. Rawls, "Kantian Constructivism in Moral Theory," 307.
39. This is presented most clearly in the case of relatively reasonable peoples, or "decent peoples" as Rawls refers to such groups in *The Law of Peoples*. These groups' limited reasonableness renders the ideal of a well-ordered society only partially attainable at the local level unless a more liberal viewpoint is endorsed. Rawls, *The Law of Peoples*, 63–78.
40. O'Neill, "Constructivism in Rawls and Kant," 353.
41. Rawls contends that arriving at a reasonable resolution to the problem arising from a plurality of conflicting views requires engaging in reflective equilibrium, which involves revising our "considered judgments" in order to arrive at a coherent set of moral judgments. But the "wide" reflective equilibrium that Rawls requires doesn't only stop at one's own judgments, but also requires developing a coherent view as a result of considering publicly shared moral conceptions as well. Rawls, "The Independence of Moral Theory," 289–90.
42. Taylor, *Reconstructing Rawls*, 239.
43. As O'Neill argues, the notion of consensus is problematic since agreeing to terms on the condition that others agree too is neither necessary nor sufficient for reasonableness. In fact, as I will suggest in this section, reasonableness in Rawls will not be adequately grounded. O'Neill, "Political Liberalism and Public Reason," 415.
44. Habermas, *The Inclusion of the Other*, 84–85.
45. Rawls specifies the conditions of stability in terms of an existing overlapping consensus and a view of moral psychology. The first condition renders the consensus actual, and the second responds to the question of whether growing up under liberal democratic institutions will produce a sense of justice in citizens. Rawls, *Political Liberalism*, 141.
46. Rawls, *Political Liberalism*, 84.
47. Rawls, *A Theory of Justice*, 251–52.
48. Rawls, *Political Liberalism*, 100–01.
49. Rawls, *Political Liberalism*, 67.

50. Rawls, "Kantian Constructivism in Moral Theory," 326.
51. Rawls, *Political Liberalism*, 85–86.
52. Rawls, "Kantian Constructivism in Moral Theory," 340.
53. Rawls, "The Independence of Moral Theory," 294.
54. Rawls, *Political Liberalism*, 68.
55. Rawls, *Political Liberalism*, 158–60.
56. Rawls, *Political Liberalism*, lviii.

References

Habermas, Jürgen. *The Inclusion of the Other: Studies in Political Theory*. Edited by Ciaran P. Cronin and Pablo De Greiff. Reprint edition. Cambridge, MA: MIT Press, 2000.

Kaufman, Alexander. "Rawls and Kantian Constructivism." *Kantian Review* 17(2) (June 2012): 227–56. https://doi.org/10.1017/S1369415412000040

Krasnoff, Larry. "How Kantian Is Constructivism?" *Kant-Studien* 90(4) (1999): 385–409. https://doi.org/10.1515/kant.1999.90.4.385

O'Neill, Onora. "Constructivism in Rawls and Kant." In *The Cambridge Companion to Rawls*, edited by Samuel Freeman, 347–67. Cambridge: Cambridge University Press, 2002. https://doi.org/10.1017/CCOL0521651670.010

O'Neill, Onora. "Political Liberalism and Public Reason: A Critical Notice of John Rawls, Political Liberalism." *The Philosophical Review* 106(3) (1997): 411–28. https://doi.org/10.2307/2998399

Rawls, John. "The Independence of Moral Theory." In *Collected Papers*, edited by Samuel Freeman, revised edition, 286–302. Cambridge, MA: Harvard University Press, 2001.

Rawls, John. "Justice as Fairness: Political Not Metaphysical." *Philosophy and Public Affairs* 14(3) (1985): 223–251.

Rawls, John. "A Kantian Conception of Equality." In *Collected Papers*, edited by Samuel Freeman, revised edition, 254–66. Cambridge, MA: Harvard University Press, 2001.

Rawls, John. "Kantian Constructivism in Moral Theory." In *Collected Papers*, edited by Samuel Freeman, revised edition, 303–58. Cambridge, MA: Harvard University Press, 2001.

Rawls, John. *The Law of Peoples: With "The Idea of Public Reason Revisited*," revised edition. Cambridge, MA: Harvard University Press, 2001.

Rawls, John. *Lectures on the History of Political Philosophy*. Edited by Samuel Freeman. Cambridge, MA: Belknap Press, 2008. https://doi.org/10.2307/j.ctvjnrtqz

Rawls, John. *Political Liberalism*, expanded edition. New York: Columbia University Press, 2005.

Rawls, John. *A Theory of Justice: Original Edition*, reissue edition. Cambridge, MA: Belknap Press, 2005.

Sen, Amartya. *The Idea of Justice*. Cambridge, MA: Belknap Press, 2011.

Taylor, Robert S. *Reconstructing Rawls: The Kantian Foundations of Justice as Fairness*. University Park: Penn State University Press, 2011.