University of Alberta

Reading Rawls's Lectures on Hobbes's *Leviathan:*
Pluralism, Stability, and Political Consensus

by

Glen Keegan Ford Gibson

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Arts

Department of Political Science

Dedication

This work is dedicated to the memory of my father, Vincent, my great friend and original source of encouragement to pursue higher education, and to my mother Francis who is a continuing source of encouragement.

Abstract

This work examines Rawls's lectures on Hobbes, which are delivered at a time when Rawls is developing his political liberalism. It especially seeks to question the foundations of Rawls's interpretation of Hobbes's *Leviathan* as presenting a core "political" doctrine based around prudential egoism. It takes issue with Rawls's moral critique, primarily by arguing that Rawls leans too heavily on viewing Hobbesian subjects as predominantly egoistic in nature. The latter approach is argued to make too little out of irreducible pluralism as a factor shaping Hobbes's argument, which Rawls only limitedly acknowledges. It is this divergence in faith over the ability of pluralistic subjects to form a political consensus over fair terms of association that shows the difference between a Hobbesian view and Rawls's own. That is, it shows the difference between the possibility for an "overlapping consensus" and Hobbes's more pessimistic outlook.

Acknowledgements

Table of Contents

Introduction

Opening discussion

This writing is occasioned by the fairly recent publication of John Rawls's collected *Lectures on the History of Political Philosophy*, and focuses on giving a careful look at Rawls's treatment of Hobbes's thought in his landmark *Leviathan.* Rawls's lectures on Hobbes begin with a number of stark statements about the latter's major work. *Leviathan* is praised by Rawls for being perhaps the greatest single work of political philosophy in English, or any language (2007, p.23). It is marked by, among other qualities, an incredible power of style, language, and intricate analysis, as well as a vividness of observation (ibid). Despite these qualities, it is, for Rawls, also a "dreaded" way of thinking about society, one which has an "overwhelming" and "dramatic" impression on our thought and feeling, made worse by the fact that it "might almost be true" (ibid).

Since Rawls has envisioned that his lectures might occasionally shed some light either on liberalism, or upon his own work, one might wonder if there is anything at stake in these articulations (2007, p.xvii). Undoubtedly Rawls's choice of Hobbes for his lectures attaches to the latter's canonical significance to the Western tradition, but what is to be made of Rawls's sometimes colorful appraisal of Hobbes? It is true that, earlier in his work, Rawls had distanced himself from Hobbes's social contract view, obliquely mentioning that it "raised

special problems," but is it really, as Rawls states in his lectures, one to be "dreaded"? Why might it "almost" (but not quite) be true?[1]

As Rawls's lectures are aimed at undergraduates, it is possible Rawls is simply aiming for stylistic effect here. Hobbes is, after all, popularly known for having held a dim view of human nature.[2] *Leviathan* also certainly invites dramatic characterizations due to its own use of colorful imagery and metaphor. This was something Hobbes's contemporaries occasionally lampooned as ironic, given the thinker's professedly sober-minded eschewal of all things fantastical.[3] It is true that in some ways the biblical Leviathan, "king over all the children of pride," is a fitting avatar for Hobbes's political philosophy – at once above man and answerable to no one (but God). Still one might take a wry pause to remember that the biblical Leviathan *also* makes an easy doppelganger for Satan. (Indeed, in Isaiah he is even marked out for divine punishment.[4]) Hobbes has undoubtedly made a rather startling choice of creature to represent his idea of the state.

A recent exposition of Rawls's later thought by Paul J. Weithman carries forward the question of Rawls's appreciation of Hobbes as a contributor to the tradition of Western thought. Weithman sees more than theatrics in Rawls's

---

[1] (1976, p.11n)

[2] One thinker who memorably touches on this is Paul J. Johnson (1987), who in his *Hobbes and the Wolfman*, argues that Hobbes's often colorful language and dramatic arguments have caused many Hobbesian scholars to over-emphasize the "selfish" or allegedly destructive side of human nature. His own view is more nuanced, and comparable to Rawls's view, in that Hobbesian man's behaviour is shaped by his circumstances, with dire circumstances making for dire men.

[3] Patricia Springborg is cited here, *Hobbes's Biblical Beasts: Leviathan and Behemoth* (1995).

[4] The quotation in this paragraph comes for Job 41:34 (The New *King James* Version). In Isaiah Leviathan is described as a "crooked Serpent" and a "dragon" 27:1. This echoes the way Satan is similarly described as a "dragon" and "Serpent" in Revelation 12:9. In Isaiah it is said that the Lord shall "punish" the Leviathan by slaying him at the end of time.

dramatic assessment. In his *Why Political Liberalism: On John Rawls's Political Turn*, Weithman goes so far as to append Hobbes to Rawls's list of "dark minds in Western thought" (2011, p.66). The reference refers to a rather elliptical passage in Rawls's *Lectures*, where Rawls singles out both Augustine and Dostoevsky as being such "dark" thinkers operating in the western tradition (2007, p.302). With Dostoevsky, this sentiment comes via Rawls's exposition of Mill. Rawls sees Mill's political philosophy as taking as a fixed point a permanent human interest in knowing the truth in all subjects. This interest is in turn supported by free institutions (ibid). Dostoyevsky's dark thought, voiced through the fearsome character of his grand inquisitor, is that knowing the truth might actually be horrible, and upon learning it we would be willing instead to support a dictatorial regime that could preserve our necessary and comforting illusions (ibid).

Hobbes is not, however, similarly identified by Rawls directly as one of his "dark minds" in his *Lectures*. Weithman as well acknowledges that Hobbes's view of human nature need not be thought of in the same manner as Dostoevsky or Augustine. [5] This follows Rawls, who, in his lectures does not seem to be arguing that Hobbes proffers an especially bleak view of human nature. [6] Yet his opening assertion that Hobbes's view offers a dreaded way of thinking about society, and makes a "dramatic impression" along those lines, would seem to

---

[5] Weithman cites a conversation with Ronald Beiner as clarifying this point for him (ibid).

[6] Rawls's brief appendix to the second lecture offers some reasons as to why this might be the case. Rawls does not read Hobbes as an egoist, nor does he think that the state of nature is caused by the great number of us having destructive passions like vainglory (2007, p.53). Rawls also wants to point out that Hobbes does indeed recognize that as individuals we may have many motivations, including those that are moral or social in nature (2007, p.46). These points are discussed in my second chapter.

suggest that Rawls would place the English thinker in the camp of the Western canon's more 'pessimistic' political thinkers. This sentiment is certainly evidenced later in Rawls's *Lectures* where Hobbes's view is said to be a "frightening" possibility. Its dramatic impact comes via the suggestion that we need not be monsters to be in "serious trouble" with regard to the state of nature, which is an ever present possibility for us even if we are by and large nice and peaceable people (2007, p.51). On Rawls's reading of Hobbes, we need not all be driven by pride and vainglory or otherwise destructive motives to end up in a state of war. We need only the lack of an effective sovereign to establish trust between citizens (ibid).

This is echoed by Weithman who sees Hobbes as putting collective action problems at the heart of social cooperation, a situation only a sovereign might stabilize. Even granting that this is the case, would it be enough to convict Hobbes of being a "dark mind" within the Western tradition? Perhaps it might be so, on the account that Hobbes like Dostoevsky's inquisitor and Augustine seem, on some level, to all advocate some form of dictatorial organization for society. The charge here is apparently one of *illiberalism* on the basis of a rejection of democracy.[7] Weithman goes further though, and broadens the question to one of the very nature of "political life" itself. On this view, Hobbes, Augustine, and other thinkers, place coercion at the heart of political organization, which is

---

[7] Hobbes of course still advocates that an effective democratic sovereignty ought to be absolutely obeyed, even if some of his observations about the nature of government seem to suggest flaws inherent in certain familiar forms of democratic organization and institutions (XIX.3).

required specifically to temper the inherent injustice of our natures. According to Weithman (2011), if Rawls could show that justice need not be "imposed" on us in this manner, he would answer such thinkers on a more profound level. He would show that a "just society" does indeed suit our nature (p.66).

With this work I take a divergent view. It is true that Rawls's critique of Hobbes does indeed take up the question of 'political life' as a question of morality. Indeed, it juxtaposes two visions of consensus: one that Rawls terms "rational" and non-moral, and another that is "cooperative" or moral. In this relation, Rawls sees Hobbes as putting forward prudential rational reasons for political consensus (individuals focus on their own self-preservation and private the means to prosperous social living) that fall short of the mark of being authentic 'moral' motivations. Political authority is instead justified in terms of the consequences that occur when subjects cannot agree to have a sovereign, which is only a little short of the suggestion that the sovereign corrects 'unjust' human nature. For Rawls, what Hobbes primarily misses here is the idea that political subjects might be motivated by a desire to cooperate based on fair terms of association for their own sake.

Rawls then gets bogged down in a number of Kantian-inspired concerns over the moral status of such a consensus in terms of human motivation. For Rawls, even if we agreed to some very authentically nice and reasonable precepts in authorizing the sovereign, at bottom the reasons that justify such agreement are simply prudential and egoistic (self-centered), and therefore disqualified from 'moral status.' On Rawls's assessment, Hobbes primarily wants us to accept that

our desire to preserve our safety and to achieve the 'goods' of peaceful social living require us to accept any political authority that makes these things possible. Hobbes then also has a number of supplementary reasonable recommendations bearing on how we ought to treat each other, which also fall out of (according to Rawls) our desire to protect our 'fundamental interests.' For Rawls these nevertheless fail to have moral significance because they are respected for the wrong (selfish) reasons.

This seems to be where Rawls lands in his critique of Hobbes, and it ends up being unfortunate (even putting aside whether Hobbes or anyone else should care about the 'moral status' of his recommendations). For one thing, what we get is a reading of Hobbesian subjects as 'prudentially egoistic,' but one that is not without ambivalence, as Rawls does not follow the thinkers he is drawing on here as far down as he could. [8] Indeed, Rawls is very careful on the question of Hobbesian human nature. He is reluctant to make Hobbesian subjects into purely rational egoists, and his lectures often seek to check the idea that Hobbesian subjects are necessarily "selfish" or egoistic intrinsically.

For Rawls, Hobbes's view of persons as egoistic is in service of a political doctrine only – 'true enough' for his purposes. On this view, only very strong self-interested reasons count for political consensus. But Rawls is less clear on why only very strong motivations such as the fear of death and other "self-focussed" goods can unite Hobbesian subjects, and why other "social motivations" are less compelling. Rawls ends up with a critique of human nature

---

[8] Particularly D.P. Gauthier (1969) in his *The Logic of Leviathan*.

by default where he suggests that these are just "very strong" and highly motivating reasons, where other reasons are not enough to prevent political subjects from acting in destructive anti-social ways. Rawls does not make very much out of his own articulations that these reasons are "universal" or that Hobbes is appealing to common ground.

Moreover, this approach also leaves standing Rawls's initial attempt to cast Hobbes's doctrine as "political" or divorceable at its core from "comprehensive" religious or philosophical claims and aimed at those things to which all persons might be expected to agree upon, whatever else they might hold. [9] Rawls asserts that Hobbes's core doctrine (his "laws of nature") is political in that it is intended to be a reasonable basis for agreement between political subjects and may even specify reasonable terms of reciprocity amongst them (even if Rawls criticizes the foundations of such principles within *Leviathan*). But what Rawls does not do is take any time to work out how this sort of "political" doctrine is importantly different from his own developing political liberalism.

It is true that Rawls does work out that his vision of "consensus" is to be a moral one, where Hobbes's vision (on Rawls's view) fails to reach that ideal. But Rawls never fully explores the difference beyond noting it and then only half-heartedly tying it to questions of human nature, before 'critiquing' its 'moral' validity. The result is a reading that seems weak on the face of Rawls's own analysis.

---

[9] Such comprehensive concerns as deep metaphysical and religious claims most certainly *are* to be found elsewhere in *Leviathan*, of course. But Rawls argues they are not essential to Hobbes's core political recommendations. This is covered in chapter two.

What Rawls misses out on here is an opportunity to satisfyingly pin down the difference in the site of 'consensus' itself; that is, exactly what political subjects are agreeing on, and what political judgement can be about from Hobbes's view to his own. Hobbes is very skeptical about the ability of political subjects to have a shared form of open 'public' judgement over how they organize and govern themselves. Hobbesian subjects must therefore be represented by the sovereign who is the standard for such public judgment. For Rawls, by contrast, political subjects can share a language of agreement and validate the terms of social cooperation on that basis. That is, pluralistic subjects can find the moral resources they require to agree on principles of justice. They are not limited to merely authorizing a sovereign to impose an external standard of judgement.

This is not strictly speaking a question of human nature, spelled out in terms of 'reasonable' versus 'egoistic' subjects. Rather it is a question of the moral makeup of political life. That is, for Rawls, fair public terms of association can take priority over competing claims, where subjects come to prioritize certain values of public life over other 'private' (though social) concerns. Citizens can form a consensus over core tenets of justice and make these stable. It is not merely a question of whether "justice" must be imposed by us from above (sovereign enforcement), or whether we can rely on the subjective moral nature of individuals to make political cooperation possible. (This is a basic problem of stability; can we rely on others to do their part without some "self-interested" motive or coercive enforcement?) We must also have a faith in the ability of citizens to work out terms of cooperation and 'do their part' with respect to a

publically shared conception of justice and not to defect out of allegiance to other socially derived concerns. (Can we have a political arrangement that also makes 'moral sense' to pluralistic citizens and retains their allegiance?)

What Rawls misses out on is drawing toward a crucial difference with Hobbes on this point. Pluralistic subjects cannot be stably unified around anything like an 'overlapping consensus' on a Hobbesian view, although they are very capable (on Rawls's reading) of social and moral concerns. This is because Hobbes handles the assumption of pluralism in a markedly different way from Rawls. For Hobbes, the politically destructive aspects of pluralism cannot be easily ameliorated. They create a similar 'problem of stability' to the one that Rawls addresses with his later writings, but with a much more radical character. That is, they foreclose the possibility of anything like "public reason," or even that Hobbes's own 'reasonable' laws of nature could be the basis for a political consensus between pluralistic subjects. Thus the sovereign must set the terms of the public world rather than these emerging as a product of a political consensus in civil society.

This work's core argument is that Rawls misses an opportunity to work through and analyze Hobbes's thought on pluralism as a problem for politics. This and not human nature simply shows the difference between what Rawls will come to call an "overlapping consensus" and the model of authorization of the sovereign from *Leviathan*. For Hobbes the only check on the divisive nature of our different conceptions of both "good" and "bad" is to surrender our power of

judgement to an all-powerful sovereign. Pluralism is a problem for Hobbes in

Leviathan in that it is the source of potential factionalism and conflict; thus, while

citizens can believe whatever they want in their own hearts, publicly they ought to

follow the sovereign's judgments (XXIX.6-7, XVIII.9). This is distinct from

Rawls's own view where "public" standards of justice and judgement, as they

relate to social life, are truly public.

Pluralistic subjects, for Rawls, can form a basic 'reasonable' language of

values and recognitions that avoids deep claims about the ends or purposes of

human life, or grounding reasons generally that go beyond what everyone might

reasonably accept. This allows for an agreement over the basic structure of society

between such subjects, including over the shape and justice of its social

institutions. As a reformist view, this allows citizens within a liberal democratic

context to formulate their own language of justice based on values they already

accept and to reposition the sovereign and public judgement in that context.

Rawlsian subjects can thus participate in validating public judgement in a way

that Hobbesian subjects cannot.

The difference between the two views, again, attaches not to questions of

human nature simply put so much as very different perspectives on what we might

hope for in the political realm as well as what can be achieved if we accept

irreducible pluralism as a fact about the social world. In very short this is the

difference between irreducible pluralism and irreducible but *reasonable* pluralism.

This is the major question of political philosophy between Rawls and Hobbes that

is missing from Rawls's lectures on *Leviathan*.

I

Short Formulation

The threads of my critique here can be woven together in the following way. Although Rawls is already engaging Hobbes using categories familiar to his later work, he fails to engage Hobbes over the shared assumption of pluralism. This is an area where Hobbes's classical thought diverges markedly from Rawls's own later vision. Using this point of analysis, Rawls would be able to more fully develop the idea of Hobbes as having a "political" dimension to his core thought, but he could also see why Hobbes stops short of being able to use the "political" as a vehicle of consensus over a set of political principles, for justice or the organization of the political world more generally.   Pluralism creates a problem of stability that only final recourse to an absolute sovereign can solve. This realization would (1) allow Rawls to make better sense of his own lectures on Hobbes and explain Hobbes's seeming reliance on predominantly egoistic motivations without having to imply that Hobbesian subjects are predominantly egoistic (something which Rawls implies he wants to avoid as he calls this assumption in Hobbes's work "political"), and (2) it would allow Rawls a better view of 'what is at stake' between Hobbes's political philosophy and what will be Rawls's later political philosophy. Differences in how each thinker treats pluralism serve to clarify the divergence between Rawls's "political" hope for an 'overlapping consensus' and Hobbes's more pessimistic view (without having to rely on an uncomplicated picture of human beings as either generally 'nice' and reasonable, or else predominantly egoistic or even selfish). With this, Rawls could

go beyond a moral critique of Hobbes's thought (the foundations of which –

predominant egoism – Rawls makes uncertain in any case) and more directly

assess Hobbes's challenge to Rawls's own 'faith' in the possibility of a moral

political consensus from the side of political philosophy, or what we might hope

for as specifically political subjects.

II

Broader themes of this work

By its end this work will briefly present the idea that Rawls misses a chance

to work out an early presentation of his own ideas of "reasonable pluralism" and

an "overlapping consensus" and compare it against an earlier compelling rival

conception to be found in Hobbes's *Leviathan.* Although Hobbes's "solution" to

the problem of political unity and representation given the assumption of

pluralism seems extreme, elements of his approach remain persuasive. Indeed, a

number of thinkers argue Rawls's idea of political consensus is also problematic,

as it presupposes a consensus on values that eliminates the element of "conflict"

in political life. This is a vision that some thinkers have gone so far as to call anti-

political. [10] But even pluralistic subjects need to find some way to be represented

as a collective body with interests in common, where they share membership in

the same political association. If some judge must ultimately arbitrate between

---

[10] Chantel Mouffe (2005) memorably described it as "political philosophy without politics" and claimed that Rawls posited one rational solution or description of how the liberal values of "liberty" and "equality" could be interpreted, when in fact there are many possible interpretations (p.52). This also echoes strongly the arguments of John Gray in his *Two Faces of Liberalism,* which will be mentioned later in this work (2000).

conflicts of value and the interpretation of such values the question inevitably arises – on what basis?

No thinker in the tradition of political philosophy is as famously wary of outright antagonism developing over conflicts of value, public judgement, expediency, or any other matter of governance, than Hobbes. Indeed, Hobbes is perhaps one of the first modern thinkers in believing that such conflicts are irreducible, meeting no final end or conclusion. This, from one side of Hobbes's political philosophy, is perhaps *the* problem of politics. Recognizing one sovereign source of authority and judgment in these matters is arguably the central gesture of our collective 'artifice' in creating the state. If in any society that standard, as represented by a supreme court, or any other centre of sovereign judgement, cannot be truly representative of most of the people, then perhaps it really only is recognized as a Leviathan – part of the art and artifice of a state that persuades by the promise of secure ordinary living on the one side, but, on the reverse, never fully veils the spectre of state power.

Rawls arguably offers what is a more hopeful view of political consensus and of pluralism. In Rawls's vision persons who are deeply divided on questions of greater values, the final ends of a well lived human life, or otherwise irreconcilable philosophical or moral views, can nevertheless converge on a consensus over certain political values (though endorsed for different reasons). These build off the "background" values of our political association itself. From these there may be the hope that society could be arranged according to a shared

vision of social justice endorsed by most if not all citizens who retain both their differences and their dignity.

This work aims to show the difference in these views, not to 'pick a side.' Its author does not know which approach offers the best view of "what we might hope for" in political life, and it has obviously chosen a rather modest medium, in critiquing Rawls's lectures on Hobbes, to approach such questions. These are background concerns, but it is nevertheless an interesting contrast to highlight in order to understand what might be at stake. It can be formulated in the following way: might we as modern pluralistic subjects authentically hope for an "overlapping consensus?" And, if not, is it then *Leviathan*, "king of the children of pride," that best answers our natures as political beings?

Chapter one: background on Rawls's writings

Introduction

It is taken to be relevant that Rawls is developing his political liberalism at the time he delivers his lectures on Hobbes, which occurred in the early nineteen eighties. As such, I begin by giving a presentation of those ideas, some of which appear in Rawls's lectures on *Leviathan*. It is not assumed that the reader is necessarily familiar with Rawls's thought, and this chapter therefore broadly aims to give an overview of Rawls's core ideas with a (non-exclusive) focus on Rawls's later writings. Certain terms will be given an explication, such as what Rawls means by a "political doctrine" as well as an "overlapping consensus" among reasonable pluralistic doctrines and Rawls's concern with "stability." Although less critical to this work's overall argument, I offer a brief overview of Rawls's early writings first.

I

Rawls's early work

Rawls's early landmark work, *A Theory of Justice,* [11] had built on Rawls's prior thought on fairness. In "Justice as Fairness" (1958), for example, fairness is described as an appropriate virtue for any social practice where participants view each other as free and equal in the enterprise. In such an arrangement fair terms can be understood as all those that participants in social cooperation could have

---

[11] Hereafter, *TJ*

agreed to had they been there to set the original rules (ibid). [12] Participants in

social cooperation, the terms of which they have endorsed and benefits enjoyed,

also have an obligation to do what is required of them when it comes their turn

(ibid). This is the "duty of fair play" whereby participants in a social practice must

reciprocate in observing the rules of the "game," which they have acknowledged

to be fair and which they have benefited from. [13]

In *TJ* Rawls carries forward this basic idea and uses a number of powerful

thought devices to model such a 'fair choice' situation. Rawls works within the

social contract tradition to do so, re-imagining the political compact as occurring

within what he calls an "original position."[14] In the OP 'contracting parties' are

perfectly equal participants in drafting the principles of justice that will

characterize the social institutions of their society (1971, p.17-22). This equality

of participants is in part guaranteed by Rawls's accompanying device, the 'veil of

ignorance.' The veil of ignorance guarantees that participants in the original

compact do not know who they will "be" in the resulting society. It follows that

they do not know what they have at stake in choosing, and they must try to make

the rules of the game they set up as fair as possible for everybody (ibid).

Participants in the original position are considered to be 'rational' actors in

the sense that they are able to make effective choices in achieving what they judge

---

[12] Here Rawls states principles of justice that more or less will later appear in *TJ*, but without the accompanying "devices" of an original position or veil of ignorance. Rawls's general concern for specifying terms of social cooperation for individuals as free and equal continues into his later "political" work where this is taken to be the clearest expression of the "fundamental ideas viewed as latent in the public political culture of a democratic society" (2005, p.175).

[13] This idea reappears in *TJ* (1971, p.343).

[14] Hereafter, OP.

to be 'good,' with the important restriction again that in this scenario no one could

know who they would be and what they would take seriously and care about in

the resulting society (1971). [15] Participants would thus have to choose rules for

social institutions that would be as fair as possible for everyone concerned. In this

scenario persons are also taken to be neither overly altruistic nor obstructively

egoistic (ibid). They pursue their own good, but the situation itself nevertheless

enforces fairness. Through all, this agreement is imagined as coercion free, and no

obvious consequences force the participants 'hand' one way or another (as one

might possibly argue about Hobbes's state of nature, for example). In this sense it

respects the idea of free association between individuals for Rawls. [16]

Rawls's original vision serves as an idealist critique of the social contract

concept itself, and Rawls offered it as a standard by which democratic societies in

particular might be judged. That is, Rawls's argument is epistemic in nature and

designed to provide philosophical knowledge of Rawls's core concept of justice

as fairness (in social practice). The kinds of principles that participants in an

original position would potentially agree to would be very fair, and thus, for

Rawls, carry an obvious moral force. With this in mind, Rawls then famously

---

[15] Rawls thought this would lead to a preference for more "primary goods," which are those things anyone could make use of to further their own 'life plan' and that society could make available, such as a guarantee of liberty (1971, p.62). Rawls acknowledges that were the veil of ignorance to be lifted some people might discover they wanted less primary goods for various reasons (religious ideals, or presumably a certain view of human interests); thus, the OP does tend to promote a certain, arguably liberal, perspective.

[16] By "free association" here I mean an agreement that *could* have been entered into by free and equal citizens. As Samuel Freeman (2007) notes, it resembles Kant's Original Contract and is hypothetical in nature. Rawls, though, reads all the social contract theorists as thinking in a similar hypothetical manner in order to devise reasonable terms of association. Rawls's vision follows these lines, but introduces moral restraints on 'rational choice' therein with the veil of ignorance device.

goes on to propose just what the kinds of principles he believes *would* be agreed

to in such a position, which he gives a lexical ordering (the first having priority

over the second).

The first and primary principle is that "each person is to have an equal right

to the most extensive basic liberty compatible with a similar liberty for others"

(1971, p.60). This principle is enshrined so as to protect the idea of free

expression as both an inherent good, and also as politically indispensable for the

vitality of a constitutional-democratic society (1971). [17] Rawls's second principle

is that inequalities would be arranged so that (a) they are to be of benefit to the

least-advantaged members of society and (b) they are attached only to offices and

positions that are open to everyone under conditions of fair equality of

opportunity (1971, p.60). These are the sorts of principles that Rawls thinks

equals in an original agreement would decide upon; that is, no one would accept

their liberty to be curtailed or inequalities to exist except where such inequalities

proved to be of benefit to them no matter where they would end up in the

resulting society. [18]

---

[17] Rawls holds a fairly egalitarian in his outlook, and he did not extend the concept of liberty to extend to private productive property as opposed to 'personal property' (1971). This is a distinction from Rawls's "libertarian" critics, such as Robert Nozick, who argue that any curtailing of the free right of exchange is an unjustified infringement of liberty and also that free exchange forms the basis for the legitimacy of any distribution of social goods (1974).

[18] The interpretation here is not uncontroversial. There are a number of alternative justifications including Rawls's "rational choice" perspective and the "maximin" principle, which states that participants in an original position would rationally seek to protect the weakest position. As Samuel Freeman (2007) notes, this idea garnered a lot of criticism over whether are not participants might be inclined to gamble, especially where other social setups might have a much greater overall "utility" or general happiness, which caused Rawls to reconsider it, though not abandon it entirely. Rawls proposes a number of other possible justifications as well including social *stability*, but Freeman particularly argues for the idea of fair terms of reciprocity where "gains" made at the expense of others would be unacceptable terms between free and equal persons (pp.188-97). I emphasise this view.

Potential problems Rawls might have had with a Hobbesian from the standpoint of his early moral philosophy then are the following: (a) Hobbes arguably derives his conclusion about authorizing the sovereign, as well as his other laws of nature, from his famous state of nature argument. On Rawls's reading, this means a state of affairs whereby the background conditions do not allow subjects to trust each other, and thus leads even regular people to anti-social distrust and even disastrous conflict. Rawls's own view of the social contract presupposes no such dire consequences, but rather an ideal "agreement" between equal subjects to fair terms of cooperation. (b) Although Rawls's proposition presupposes "rational" actors, its content is meant to represent fair principles of cooperation that are endorsed for their own sake. Rawls's view takes fairness as a good all on its own, whereas Rawls thinks Hobbes tries to justify the reasonable content of his theories in terms of the rational (as I will note later, primarily "self-preservation"). These concerns also are reflected in Rawls's later work, but Rawls orients these around what he will term "political" concerns. With that said, I will now turn to Rawls's later work.

II

Rawls's later work

Rawls's work in *TJ* received a lot of critical analysis and commentary. Those engaged in such criticism received a generous response in turn, with Rawls

spending a lot of time responding to the reaction his work generated. [19] It is

nevertheless hard to gauge what role this criticism (particularly the charge that

Rawls's system did not take pluralism seriously) played in leading Rawls to re-

envision the place of his core ideas, or whether Rawls found compelling internal

reasons to do so all on his own. Indeed, as Weithman (2011) notes, Rawls puts

forward his own explanations, which are ostensibly related to issues of stability,

though there are different interpretations of his views.[20] In any case, with Rawls's

later work, particularly *PL* and *Justice as Fairness: a restatement*,[21] Rawls began

to re-orient his basic conception of justice in a way that, to quote the title of an

import article in this turn, was *political* not *metaphysical* (1985).[22]

Following this line, *PL* takes as its framing problem the following question:

"how is it possible for there to exist over time a just and stable society of free and

---

[19] In *Rawls: a theory of justice and its critics,* Chandran Kukathas and Philip Pettit mention three major strains of criticism. The first is the libertarian line of criticism exemplified by Robert Nozick, with the second being a Communitarian critique of Rawls's Liberalism from thinkers such as Alasdair MacIntyre, Michael Sandel, Charles Taylor and Michael Walzer. The third is Rawls's own "self-critique," which they argue shifted Rawls from a Kantian foundation to something closer to a Hegelian one (1990). Pettit and Kukathras focus on the series of articles that Rawls would publish in the 1980s which would lead up to Rawls's thought in *Political Liberalism*. Critics such as Lukes (1978), Pettit (1974), David Miller (1976) and Milton Fisk (1978), criticized Rawls for presenting a theory of justice that was essentially Western and individualistic, which Petit and Kukathras argue shaped Rawls's "political turn" . Communitarian critics, especially Sandel, criticized Rawls's allegedly "metaphysical" conception of personal identity, and this also spurred Rawls on in this respect. Pettit and Kukathras argue these criticisms led Rawls to more directly articulate his position as *political* and not comprehensively moral or metaphysical – appropriate to be accepted by pluralistic liberal individuals, from a variety of positions (1990, pp.121-135).

[20] Paul Weithman mentions two possibilities here. The first, which he rejects as subtly incorrect, is what he calls the "public basis view," which is that the argument in *TJ* was too controversial to be accepted in a pluralistic society and thus required Rawls to reframe his work as "political" (2011, p.19). This is roughly the view mentioned in footnote (9). Weithman however takes Rawls at his word that he was dissatisfied with the account of stability in part III of *TJ* Stability will play an important part in this work.

[21] Hereafter, *JF*.

[22] All references to Rawls's articles are given their original dates, but these have been taken from Rawls's *Collected Papers* (1999).

equal citizens, who remain profoundly divided by reasonable religious,

philosophical, and moral doctrines?" (2005, p.4). What Rawls sought to do here

was to recast his justice as fairness in terms of something that could be acceptable

as a basis of agreement between modern pluralistic subjects who, Rawls

acknowledged, would not all share a similar worldview. This sort of an approach

would entail repositioning Rawls's ideas in terms of what Rawls would call a

"political doctrine," as opposed to one that required the individual to accept

certain deep metaphysical or comprehensively normative claims. By using the

idea of the "political," Rawls meant to suggest that his vision of justice could

form the basis of agreement between individuals whose deeply comprehensive

moral and philosophical views, however grounded, would otherwise be

irreconcilable.

Familiar to this turn in Rawls's thought are a number of key terms that will

need to be explicated in order to give this work's overall analysis, which takes

these concepts to be important, some context. This section therefore offers a

closer examination of the distinction between "political" and "comprehensive"

doctrines, which has been raised already. It will also explain Rawls's notion of an

"overlapping consensus" on what Rawls terms a "free-standing" political view

like justice as fairness, in contrast to a "mere *modus vivendi*" (2003, p.192). A few

words will also be said about the role of public discourse, and how Rawls

envisions "reasonable pluralism" and the role of stability in his thought. This will

include some words on the place of the "moral powers" Rawls supposes we must

imagine citizens as having in order to make his ideas work, which include the moral power of being "reasonable."

I begin first by saying a few words about how Rawls repositions his core concepts in his later work. In both *PL* and *JF*, the original position and veil of ignorance are presented more straightforwardly by Rawls as political modelling devices for how citizens *might* go about choosing principles of justice for a society. Here Rawls (2005) is clearer about contextualizing his devices as specifically appropriate for political subjects who *already* view each other as both free and equal and recognize themselves as participants in a scheme of social cooperation over time. These devices no longer provide something like a perfectly objective epistemic view point from which we can deeply know what is "fair" in this context.[23] Rather, Rawls proposes these as a way pluralistic people, who nevertheless share a commitment to treating each other as equals according to "fair principles of reciprocity," might think about designing the basic structure of their society.

---

[23] By epistemic device I mean simply that it gives a kind of knowledge about what "just" or fair terms of association might be. In *his Justice and the Social Contract*, Samuel Freeman says these give a view into the "'correct concept' (quoting Rawls) of the principles that define a just constitution," which is what free and rational individuals would have agreed to if they were "equally situated" (2007, p.145). Rawls received some criticism for his particularly strong statement in *TJ* that he had provided an "Archimedean point," from which the basic structure of society could be judged (1971, p.584). Examples of such criticism can be found in the communitarian critique of Rawls. In "Justice and the Good" (1984), Michael Sandel specifically raises the term in criticising Rawls for envisioning a universal vantage point for justice that abstracted from persons' individual identities and rootedness in certain communities and traditions.

In the OP persons are now imagined as if purely rational representatives of a theoretical citizen, and are out for his or her own good. [24] They then decide how to arrange the basic structure of that society to best suit that individual. (Here the "basic structure" means a polity's constitution, and how the laws might bare upon social institutions like the economy or even the family [2003, p.10-11].) The veil of ignorance again prevents such "representatives" from knowing any details about what position the individual they represent will occupy in the resulting society, and their choice under such conditions is said to model "reasonability" (2005, p.304-10). Indeed, what these devices do overall is to *model* two separate moral powers of citizens, namely their "rationality" and "reasonability." (They provide roughly a sense of what is good limited by a sense of what is fair.)

That basic framework would cover the constitution and the laws that apply to the basic structure of society (which is actually quite a broad range of territory). Although we would want to imagine the most *basic* principles of fairness from a position of 'blindness' to difference, Rawls imagines our applying these in a step by step process where we gradually "lift the veil" and try to apply our ideals in context – to a particular society with particular identities and a particular history. [25] Although this is all very similar to *TJ*, Rawls wants us to see this whole process

---

[24] In his later work Rawls (2003) is clearer about stating that in the OP one imagines oneself as a *representative* of a theoretical citizen in society, rather than as a party directly representing oneself. This turn is curious, and it is not entirely clear why Rawls makes it. Joshua Cohen argues that imagining oneself in the role of "representative" clarifies one's position as occupying a kind of public role and trying to work out the most responsible conception of justice possible for the common good amongst equals (pp.101-2). This is something Cohen sees operating in *TJ*, but the language more clearly points in this direction in *JF*.

[25] This is a four stage process; with each stage progressively more information is required to apply the principles of justice in context. The stages are the following: adopting principles of justice, constitutional convention, legislative process, and finally administration and interpretation by the judiciary (2003).

now as a series of practical modelling devices that individuals holding otherwise irreconcilable views (religious, philosophical, moral) might nevertheless use to specify fair terms of reciprocity between themselves, which they would all then morally affirm from their own particular standpoints. [26]

That is the basic outline. A closer analysis of a few Rawls's key terms and concepts, and the role they play in Rawls's overall view, will now be offered. For instance, what does Rawls mean by saying that his vision would be a "political" doctrine rather than "comprehensive"? To say that a doctrine moves within the "category of the political" is, on this view, to say its arguments do not depend on any comprehensive religious, moral, or philosophical outlook (2005). Comprehensive doctrines, conversely, may be known by their tendency to apply to all areas of life; that is, Rawls tell us that they can inform what our final ends might be as persons, or how we ought to act in very many situations, and they may also specify ideals of personal conduct and notions like friendship (2005). [27]

The philosophical doctrine of utility is a relevant example. One might crudely understand this doctrine as the dictum that the consequences of our actions should aim towards producing as much "good" or "happiness" as is

---

[26] This view has left out two aspects for the purposes of brevity. Such pluralistic citizens operate from what Rawls refers to as ideas "latent in a public political culture of a democratic society" (1988, p.252). Rawls's method herein is "political constructivism," which in the context of Rawls's later works means starting with the fixed intuitions of a democratic society and using a particular procedure to try to construct a view that best represents those intuitions (Hedrick, 2010, pp.52-3). Such a view of justice as we are able to formulate is then always checked against a "reflective equilibrium," which means our collective considered judgments regarding political justice (2005, p.28).

[27] According to Samuel Freeman these include conceptions of "what's of value in life and gives life its meaning" as well as metaphysical doctrines that bear on the "nature of reality" and "epistemological doctrines" that have to do with "the possibility and conditions of human knowledge" (2007, p.332).

possible across however many subjects. Such a principle, for Rawls, could easily govern political decisions, but equally so all aspects of choice in an individual's life. What is more, someone who did not recognize the same fundamental principles might be reluctant to accept any arguments that are made to turn on its key premise. Part of what Rawls calls "modern conditions" is that it is not reasonable to expect that all citizens will share such a deeply grounded outlook (2005, p.36).

A "political doctrine" for Rawls makes no claims that someone who held a deeply rooted but "reasonable" religious or philosophical view could not accept. Such doctrines are meant to be affirmed from a diversity of possible deeply-rooted groundings, but are not argued from those particular vantage points. One might endorse the conclusions of something like justice as fairness on the grounds of utility, for instance, but one would not attempt to justify this view on that account to a devout religious believer, or anyone else, who did not recognize the grounds of "utility" as a political reason (2003). The justification provided would instead refer to the shared recognition of the conception itself, which in turn would be endorsed for other deeply rooted reasons. This is what Rawls means by saying that such a view would be "free-standing." It would not rely on any one comprehensive view for justification or for the moral allegiance that citizens are able to form to it. Instead it could be endorsed from any number of deeply held comprehensive view points or else of its own account (2003). [28]

---

[28] Samuel Freeman (2007) emphasizes that a "free standing" political view is part of a public democratic culture and separate from any one comprehensive grounding. Such a "political" view thus must stand on its own even though any number of comprehensive views might find reasons to

It is debateable whether Rawls makes enough out his reliance on the fact of "reasonable pluralism" to make this work. Rawls assumes that even though modern constitutional-democratic society is irreconcilably divided by pluralistic views of the "good," and other deeply held conceptions, nevertheless citizens in a liberal constitutional democracy share a commitment to the values of liberty, the equality of each, and have some kind of a recognition of one another as being participants (as I have said) in social cooperation over time. Rawls thinks that this last recognition, as well as the recognition of equality, suggests at least some desire to ensure the "rules" of such cooperation are fair. In this respect Rawls continues his central concern for fairness that animated his work since even before *TJ*, but this is of course one of Rawls's most contestable claims. [29]

One important aspect of this view as well is Rawls's "political" conception of the person. That is, for the purposes of his doctrine Rawls asks that we must imagine political subjects as having what he calls "two moral powers." The first is a sense of rationality, which Rawls gives a fairly simple (not particularly metaphysically loaded) definition. "Rationality" in this case is simply to say that political subjects can know what they want, can form a "life plan" of sorts around their most important goals, and they can pursue their ends more or less effectively

---

oppose them. Such comprehensive doctrines might also find reasons to endorse the free standing view, and when we have enough of these we have an overlapping consensus.

[29] As Jeremy Waldron (1999) points out, even if one did not assume that particular and incommensurable visions of justice did not flow directly from our 'comprehensive' views of the good, which is an open question, notions of justice in society still greatly vary. Rawls seems to think that in constitutional democratic culture "fundamentals" about justice, core principles and ideas, are shared by citizens, and he does not seem to allow for much room over the idea that interpretations of justice – however grounded – might be very divergent, right down to the essentials of how to understand "freedom" and "equality" as political values, and what accepting them means we should try to advocate for in politics.

(2005, pp.176-7). Secondly, and crucially, citizens also have the moral power of being "reasonable." This to say that citizens can have a "principle dependent" desire to act from fair principles of reciprocity for their own sake. The second "power" must take priority over the first for Rawls, which is to say that a desire to propose fair terms of association, and to act fairly generally, limits our pursuit of our own "good" (2005, pp.171-211).

Rawls thinks it is not overly idealistic to suggest that human beings can act in ways consonant with political fairness. But he does not claim to be making a definitive statement about the way human beings 'are' in reality with this; rather, what Rawls outlines is what he calls a "political" conception of the person, as political subjects as "designers" of principles of justice would require these two faculties to 'make it work.' [30] This is an aspirational view, then, and Rawls even goes on to imply that any reading of human nature that excludes our using our "moral powers" in this way would be something like a species of 'bad faith' over the possibility of a society well-ordered by principles of justice. [31] On this view, for subjects to have the kind of moral natures they require, Rawls thinks that they cannot derive their sense of reasonableness from their sense of self-interest or "rationality" – a concern familiar to his earlier work, now given a more distinctly 'political' gloss.

---

[30] Rawls does not use the exact term here, but this is the idea – see also *JF* (2003, p.19).
[31] Weithman argues that Rawls's political conception of the person is aspirational. We understand collectively our ability to be the sort of citizens that justice as fairness requires of us. This is our "free and equal self-conception" we must live up to in order to vindicate or faith in the possibility in a more just world (2011, p.11-4).

III

Public reason, 'overlapping-consensus,' and *modus vivendi*

So far this work has sketched a general framework of Rawls's later thought. The political compact is there imagined as taking place between citizens who are deeply divided by comprehensive doctrines, but citizens holding such doctrines are nevertheless capable of forming a "political conception" between themselves. Here subjects who have Rawls's moral powers may come to specify fair terms of reciprocity, whereby they pursue their own good but under the limiting conditions of fair terms of association. Rawls's justice as fairness is presented as an obvious candidate for such a compact. While it is still obviously an "idealist" standard, Rawls's vision is presented not as metaphysical but rather as a particularly resonant expression of what Rawls thinks are the core values of constitutional liberal democracy.

Pluralistic citizens' commitment to reciprocity and treating each other as equals means that, on fundamental matters, Rawls's political subjects also tend to deal with each other only by raising propositions that they think others might reasonably accept. They thus do not speak from a position of their comprehensive viewpoints, as in doing so they would be asking other citizens' reasons, or authorities, that they do not share. Rawls's pluralistic citizens also generally avoid proposing anything that it would obviously not be reasonable to expect others to agree with according to the shared values of the political conception (2005). In doing so such individuals are recognizing what Rawls terms their 'duty of

civility.' [32] Rawls further implies that this is how subjects in a pluralistic

democratic society ought to treat each other.

Citizens in this case tend to speak only in the shared language of the

political ideals of the society itself, and they avoid the language or claims of their

comprehensive view points, as well as raising anything else that is obviously

unreasonable or not in accord with the 'duty of civility.' This is roughly what

Rawls terms 'public reason,' and this is how pluralistic citizens in constitutional

democracies represent themselves and argue publically toward principles of

justice. [33] It is the standard of public justification, and it obviously depends on a

kind of 'overlapping-consensus' between pluralistic citizens over what the

"public" values of justice will be. [34]

For Rawls, pluralistic liberal subjects can then nevertheless reach a moral

consensus on matters of justice. Such a consensus would be what Rawls terms an

---

[32] This is the moral – not legal – duty that citizens have when it comes to fundamental questions. One must explain how the "principles and policies" one advocates and votes for can be supported by the political values of public reason; this includes also a willingness to listen to others and to make accommodations where appropriate (2005, p.217). Rawls expands this idea with the "wide view" of public reason, where citizens might introduce reasons from their comprehensive doctrines, provided that they justify their reasons in terms of the shared political association somewhere "in due course" (2005, l-lii).

[33] As Kent Greenawalt (1999) notes in "On Public Reason," Rawls's chapter on public reason is his most original contribution to *PL*. According to the view Greenawalt presents, public reason makes the idea of "reasonable pluralism" need to be a little less demanding; that is, a full overlapping consensus on principles of justice is not required, but rather public reason can proceed on a much thinner consensus over "constitutional essentials" leaving more room for disputes about basic justice (1999, pp.250-1).

[34] Rawls offers the example of the American Supreme Court here. In that forum one (in theory at least) argues only from the constitutional essentials of the society – presumably accepted by most citizens – or else one argues from the core principles or values represented by those essentials. Rawls gives as one, somewhat controversial, example the problem of abortion (2005, p.243f). On this view, one might use "public reason" to work through such an issue at the highest level by referencing society's reverence for life and desire to respect it, but, on the other hand, one would need to weight this against the importance of the liberty, freedom, and bodily integrity of female citizens. What one avoids in a sufficiently pluralistic society is raising deeply comprehensive religious or philosophical views as 'political' reasons.

'overlapping consensus' between comprehensive doctrines, which is a key idea for this writing. According to Rawls, persons holding different comprehensive doctrines can agree to a conception of justice that they affirm on moral grounds. In *PL* an 'overlapping consensus' is defined primarily as a question of stability; that is, whether or not, after we would have chosen principles of justice, our democratic pluralistic society could make it stable by affirming it (2005). In *JF* Rawls also reiterates that such a conception is to be first (a) freestanding, as in not based off of any comprehensive view, and secondly (b) "political in the right way," which is to say that it is nevertheless affirmed on moral grounds, becoming the basis for public justification, and it is not the outcome of a mere compromise or bargaining between existing political interests (2003, pp.187-8).

The view of politics as political in the "wrong" way – a process of bargaining or a coincidental overlap in interests – is what Rawls refers to as a "*modus vivendi.*" [35] Its chief problem, according to Rawls, is that the terms of association are not deeply endorsed by the agreeing parties and may be defected from at any time when one party feels it can realize the advantage (2003). This is

---

[35] *Modus vivendi*, roughly translates as a "way of living." It is a term for an agreement between parties who otherwise might disagree but who wish to go on about with life and so must find some basis for compromise. *Modus vivendi* is also a term for a treaty in international relations between nations that is meant to be replaced by a more thorough agreement later. Rawls often compares the *modus vivendi* generally to its application in international relations. In *PL*, for instance, Rawls describes such an arrangement as being like a "treaty between states whose national aims and interests put them at odds" (2005, p.147). This is echoed by Thomas Pogge who gives a good expansion on the idea of a *modus vivendi* as it applies to international relations in "International Relations as a *Modus Vivendi*" (1987). For Pogge, *modus vivendi* arrangements are characterized by a careful balancing of advantage between parties, and as such are unstable; indeed, they may even need to be renegotiated should the equilibrium shift significantly changing the balance of advantages. Pogge argued for an overlapping consensus in international relations, a more ambitious approach than Rawls will take in his *Law of the Peoples* (1999).

because it is like a settlement between parties who do not see themselves as being in social union, and exists primarily for the purposes of mutual advantage or self-interest. The paradigmatic case of such an arrangement for Rawls is early religious toleration in Europe, which must have begun as a *modus vivendi*. Here toleration proceeded because peace was preferable to war but for a while may have been abandoned if one side thought it could reach the goal of dominating society by force.

Rawls's preferred moral conception requires subjects to act from a position of fairness, as I have noted. It takes the 'political conception of the person' as allowing that such a view, as opposed to the '*modus vivendi*' is possible. This is a feasible but aspirational goal for Rawls. Indeed, Rawls phrases this memorably as a "realistic utopia" in his *Law of the Peoples*,[36] which echoes Rousseau's well known formulation in taking "men as they are, laws as they can be" (1999, p.7). For Rawls, this takes our "moral natures … within a framework of political and social institutions … [with] laws as they should, or ought, to be" (ibid). Part of "realistic" however meant not just establishing that people could act "morally," but also from the more 'political' side of things that an agreement on justice could be stable.

---

[36] Hereafter *LP*.

IV

Rawls on the question of stability

Stability becomes a question for Rawls in his later work and with respect to his vision of a "realistic utopia." With modern conditions, Rawls familiarly tells us, no one comprehensive view – philosophical or religious – can get us to social unity. [37] Stability must depend on an overlapping consensus of reasonable doctrines on a reasonable conception of justice to realize a pluralistic social union, or else it would be, again, "merely" a *modus vivendi* (1999, p.16). This concern for stability appeared in Rawls's earlier work as a question more directly of moral philosophy. That is, in Rawls's earlier work the question is whether an effective sense of justice could be decisive over our natural temptation to be self-interested and defect from doing what is necessary (think the 'duty of fair play') to maintain a system of justice. Rawls gives as one example the desire to 'cheat' on our taxes, where our moral motivations springing from a sense of justice must override such temptations, which may advance some good for us, or else the conception will not be intrinsically stable (1971, pp.336-7).

The question here then, to echo Weithman particularly, is whether stability could be inherent or must be imposed (2011, p.50). Rawls's arguments for stability in this case, as well as Weithman's assessment of them, are too complicated to fully reproduce here. However, a critical feature is that our

---

[37] Such a unity would be a true community, which large pluralistic societies cannot be for Rawls (2005). John Gray (2000) is one thinker who argues that this points to a strong connection between Rawls and Hobbes's thought.

temptation to defect is increased when we think others are likely to do so as well, which Rawls goes so far as to give game theoretic presentation (1971). As we do not want to be taken advantage of, so we must need some assurance that others will also do their part to maintain the just social scheme. Tellingly, Rawls invokes Hobbes here as an example of "imposed stability" where the sovereign stabilizes the 'decision table' by making it suitably undesirable to defect from doing what one is supposed to do. Social cooperation is what people must want anyway, but they require the sovereign to assure that their interactions proceed in a stable fashion. The sovereign allows them not to fear that others will not do their part, and hence shields them from exploitation (1971, p.240).

Rawls does not think that a society can get by without things such as penalties for dodging one's taxes (ibid). But with the idea of inherent stability, he meant that justice as fairness could generate its own supports in a significant way. [38] In short, it is known that most people in society share a sense of justice and recognition of the principles of social cooperation, and this gives assurance that others have allegiance to the vision of justice shared in society. Thus such a vision generates its own support and becomes what Rawls terms "inherently stable" (1971, pp.436, 498). In this vision minimal coercion is required, and thus justice is not "imposed" on us in the manner that Rawls believed Hobbes prescribed.

Aspects of this idea carry forward to Rawls's later work. Citizens must still acquire a sense of justice, and Rawls still argues that justice as fairness must

generate its own supports. That is, it is argued that justice as fairness could be in inherently stable once just institutions were in place (2005). However, Rawls found problems with this part of his work in *TJ*, which Weithman argues spurred him on particularly to develop his later thought in *PL*. The problem, greatly simplified, is this: *TJ* assumed that everyone would share the same partially comprehensive doctrine in respect of Rawls's justice as fairness, which could specify abstract principles that bear decisively on individual autonomy, friendship, and other fairly 'thick' comprehensive concerns. Rawls realized this was unrealistic, not only because society is pluralistic about such views but also because the very free institutions Rawls recommended with his justice as fairness could be shown themselves to generate such pluralism over time (Weithman, 2011).

According to Weithman (2011) this became a problem for the "publicity condition," which is that everyone needs to be assured the most others in society follow the same public conception of justice (pp.27, 203). How could they if they did not share the same comprehensive viewpoint? Rawls's answer was to recast "persons" as "citizens," and to make such values more specifically part of the public domain (2011, pp.357-8). Now instead of the concept of the person, there is the ideal of the public citizen and the values of the shared political association. An "overlapping consensus" means that it is not assumed that the values of the political association rely on any one comprehensive view; rather, by acting in accord with public reason and showing other signs of allegiance to the public

conception of justice citizens give each other assurance that they will act as they need to in order to keep the public conception of justice viable.

It is understood however that the ideals of citizenship and the public conception of justice retain the allegiance of most citizens, and that those citizens must find their own ways to reconcile those ideals with their own comprehensive doctrines. As Weithman notes, for those holding liberal views this might be fairly easy, but for those that are differently rooted (such in "traditional faith") this may be considerably more difficult (2011, p.300). Rawls's sentiment that "a reasonable and effective political conception may bend comprehensive doctrines toward itself," is worth remembering here (2005, p. 246). [39] An overlapping consensus on a freestanding political view does not presuppose a fortunate coincidence of values, or a "minimal best" between those who differ. Rather it suggests that "reasonable" comprehensive views have found a way to mutually endorse a shared public view of justice amongst citizens.

For Rawls this means that the view can be morally endorsed and citizens can feel assured that other citizens will do their part. This view allows Rawls to vindicate the possibility for a moral, as opposed to self-interested, view of politics and justice. It does this by giving assurance that the public moral conception is deeply endorsed because it lines up with pluralistic citizens more deeply held "comprehensive" motivations. When citizens do their part in the public arena, we

---

[39] It is also important to note that Rawls (2003) thinks most comprehensive doctrines are only "partially comprehensive" and thus allow some room for other sorts of values to take root. There is a degree of what Rawls refers to as "slippage," which means that many comprehensive views do not cover so much that justice as fairness is excluded by troublesome value conflicts (p.193).

can feel assured that they endorse the situation, and thus we are less temped (as citizens) to defect, for our own self-interested reasons, from our responsibilities in this regard.

This is still close to the line of thinking from Rawls's early work where people need some assurance that other members of society will have a "congruence" between their notions of what is "right" to do and their sense of what is "good" for them. But there is a more obvious political question in this beyond whether persons will decide to act according to moral obligation rather than self-interest, which Weithman notes that Rawls raises as well. Rawls does this where he weighs the possibility that the "values" of the political association need to outweigh the temptation to use political power to enforce such *other* values as spring from the various comprehensive doctrines in society (2003, p.89).

Indeed, our pluralistic "comprehensive" views may give us rival allegiances that cause us to act in ways that are not obviously "reasonable" to other citizens and that do not give them reason to think we will honor the public conception of justice. How can Rawls address this dynamic? It is true that Rawls handles this somewhat by starting us off at a point where we ostensibly already have "reasonable" comprehensive doctrines. Yet Rawls could not deny that even in the United States, the society he has most in mind, one might very plausibly doubt that this is entirely so. Here it is important to note that Rawls also argues "justice as fairness" *could* generate its own support, including by generating reasonable pluralism, only *after* its implementation, as Weithman observes (2011, p.346). Hence the fact of not-so-reasonable pluralism is not taken to knock down Rawls's

argument, as Rawls's ideal has an aspirational character that does not necessarily assume that the support for it is entirely existent (ibid). [40]

There are two problems of stability in Rawls's thought then. The first is that a kind of "free riding" or self-interested advantage seeking might take precedence over fair terms of social cooperation. This is so because either (a) this reflects the way human beings really are or (b) we cannot have enough assurance that others will to their part when it comes to be their turn. Rawls seems to draw on both of these with his lectures on Hobbes, as we will see. Rawls argues toward (a) where Hobbes is said to present reasonable even potentially moral terms of reciprocity only in terms of self-interest, and Rawls argues toward (b) where Hobbes is said to propose that a sovereign is required to stabilize these terms by effectively changing what it is rational to do. Thus the Hobbesian model "imposes" the stability of its principles and does not rely on the moral natures of subjects.

This concern is roughly in line with Rawls's earlier work. But Rawls is also developing his political liberalism at this time, and this comes through as well. That is, Rawls ends up giving the impression that Hobbes deals only with the problem of self-interested defections from cooperation by giving subjects a clear view of their fundamental interests and proposing a way to secure those interests. Lost here, however, is a home for all of Rawls's nuanced articulations about

---

[40] Rawls (2005) confuses the issue by constantly referring to the "fact" of reasonable pluralism, while in *PL* he explicitly states of an overlapping consensus that it "consist(s) of reasonable comprehensive doctrines likely to persist and gain adherents over time *within* a just basic structure (as the political conception defines it [p.141])." The emphasis here is mine. Undoubtedly there is a little circularity here between Rawls's need for "reasonable pluralism" to allow for something like justice as fairness, and the view that justice as fairness generates its own "supports" by fostering reasonable pluralism *after* it is implemented.

Hobbesian subjects' more complex "social" interests and also, crucially, an explanation for what Rawls is trying to do by putting forward the idea that Hobbes's core doctrine is a "political" view.

This latter train of thought leads to a second problem of stability. This is not the problem of defection from a scheme of social cooperation for "self-interested" reasons, or 'free-riderism,' as a question of the presence or efficacy of moral motivations. Rather it is the more directly *political* problem of consensus where deep pluralism creates values and conceptions that cause citizens to challenge the conception of justice out of rival value allegiances. Rawls does not seem to be far enough into his development of *PL*'s ideas, or else he feels it inappropriate for his lectures, to compare his overlapping consensus with Hobbes's "absolute" sovereign on this point.

With that said, this work's next chapter aims to work out just exactly what Rawls does have to say about Hobbes's doctrine by giving a presentation of Rawls's lectures on Hobbes's *Leviathan.*

Introduction

With an outline of Rawls's theory complete, I will now go into Rawls's presentation of Hobbes's thought in *Leviathan*. I will show Rawls's approach to Hobbes as a "political" thinker, with Rawls arguing that Hobbes advances core claims that do not depend on a deeply religious or metaphysical grounding. Although Rawls does this, he does not use this idea for his overall analysis. That is Rawls notes *Leviathan's* approach here but does not tell us why Hobbes structures his core political claims in his manner. (Rawls does not tell us if Hobbes tries to solve what I have termed "the second problem of stability," attaching to irreducible pluralism, with this.)

Instead, Rawls tells us that Hobbes's "political doctrine" attaches to prudential egoism. Rawls notes that Hobbes seeks a "universal" kind of common ground between political subjects with this, but he more often stresses that Hobbesian subjects prudentially seek to protect their fundamental self-focussed interests (as they do in a state of nature) because these are typically the most compelling or important sorts of political reasons. This is so, even though Rawls is careful regarding Hobbesian human nature. That is, Rawls notes that Hobbesian subjects are politically rather than deeply egoistic, and they are still capable of forming many social interests and concerns.

Ultimately, Rawls presents Hobbesian subjects as *'predominantly'* egoistic, with fundamentally self-focussed motivations being the strongest kinds of political reasons. This creates a kind of problem of 'stability' that Rawls thinks Hobbes meets by suggesting a kind of "mutual advantage" solution, which Rawls then critiques on the grounds of concerns appropriate to moral philosophy. The larger political dimension of stability, which takes into account complex pluralistic value concerns, is not nearly emphasized to the same degree. My next section begins with the question of Hobbes's "political" form of argument in Rawls's lectures.

I

Hobbes's "political" form of argument

I begin this chapter with some remarks on the "political" aspect of Rawls's analysis. Rawls starts his lectures by noting that Hobbes's contemporaneous critics singled his work out for, among other qualities, its alleged psychological and ethical egoism, its atheistic materialism, an alleged moral relativism and subjectivism, as well for the suggestion that power alone legitimates political authority (2007, p.25). Rawls notes these interpretations but is clear at the outset that he does not necessarily agree with them. Instead, Rawls begins by presenting the idea that, at its core, Hobbes's *Leviathan* offers principles of reciprocity that specify reasonable terms of association between equals in society (2007). Such terms are reciprocal because others are expected perform likewise, and they are (intuitively) reasonable because they could serve as fair basis for cooperation.

Rawls will later argue that their ultimate basis, however, is rational "mutual advantage," which he will find to be problematic (2007, p.54-5).

Nevertheless, Hobbes's core political recommendations are said to represent a "political" doctrine in their content, as *Leviathan* relies on neither Hobbes's "theological assumptions" nor his materialist metaphysics, according to Rawls. Hobbes's thought can instead be taken as a "secular political and moral system" in this very much apart from these sources (2007, p.26-9). Rawls (2007) explicitly refers to this system as a "political doctrine" in rejecting the idea that Hobbes is a "thorough going" egoist; rather, *Leviathan*'s core doctrine is said by Rawls to stress "certain aspects" of human life and to model human behavior that influences humans in "political situations" (p.51). Thus, Hobbes's view might be detached from more "comprehensive" concerns in terms of its core political principles.

With regard to theological assumptions, for one, Rawls argues that Hobbes's "political and moral system" is left fully intact without any religious grounding (2007, p.30). Such grounding may still be important for Hobbes, and in some way lend coherence to the ideas of natural right, the state of nature, and similar. [41] But, whether or not Hobbes's laws of nature – how we should act if we

---

[41] Leo Strauss (1963) argues that Hobbes's method begins with skepticism. In this case Strauss was referring to moral skepticism, where *Leviathan*'s "science" starts by try to find what moral premises might survive our doubts. It seeks an irrefutable premise to then begin reasoning from (similar to Descartes's approach). According to Strauss, Hobbes hit upon the imperative of self-preservation, and this is the "natural right" from which obligation is derived (p.13). However, although the desire to preserve their own lives may be the most compelling motivation humans have, certain religious notions – superstitions and the like – might overwhelm this passion; that is, fear of powers "invisible" might become more compelling for us (p.26). Strauss notes that Hobbes's religious arguments aim to foreclose this possibility. In this last aspect Rawls seems to be in agreement. Hobbes's religious or Christian arguments may not be essential to the content of

wish to have civil peace – are reasonable principles or commands of God, Rawls

observes, their *content* remains the same. What one ought to do does not change,

no matter what view one takes of Hobbes's laws (2007). On this reading, Hobbes

presents the laws of nature as right principles of reason. They are what we ought

to do to achieve peace, and they always bind *in foro interno* to a right desire that

they take place; however, *in foro externo*, or as it comes to what we do in the

world, we might not always have reason to act on them.[42] That these might be

commands of God and that we might be punished for not following them is a

secondary concern. It does not change the content or persuasive force they should

have for us in any case (2007).

Rawls likewise passes over Hobbes's metaphysics of "matter in motion,"

which appears early in Hobbes's *Leviathan*. Rawls again sees this foundation as at

best supplementary and excludes these as being necessary grounds for accepting

Hobbes's political doctrine (2007, p.29-30). Instead both common sense

observation and Hobbes's reading of the classics – Thucydides, Aristotle, and

Plato – are taken to be the major basis for Hobbes's account of human nature.

Rawls does however place some emphasis on Hobbes's so called 'reductive

---

Hobbes's core recommendations, for Rawls, but Hobbes may have needed to provide them as a helpful supplement to shield his arguments from a religious line of criticism (2007).

[42] The laws of nature dictate that we should act in such ways as accord with civil peace but only when others are so willing. According to *Leviathan*, we might have a valid reason to not do so, where we suspect, in the case of laying down our right to all things, others may not act reciprocally. Since the laws of nature derive from the first and most fundamental right to preserve our own lives, it follows that we are not bound to offer terms of peace where these might reasonably makes us vulnerable to danger (XIV. 4). Rawls chooses not to read Hobbes's laws of nature as suggesting that an "*in foro interno*" desire that they take place attaches to any sort of firm moral *obligation* on us. For Rawls they merely have practical content as principles of right reason when there is a Sovereign. (Rawls, then, excludes the idea that these might be read as deontological principles, as in Taylor [1965], or the idea that they may have commanding force as specifically as commands of God, as in Warrender [1957]).

method' (imagining the State as if dissolved or taken apart) as both the basis for his "philosophical knowledge" and perhaps our way to best understand the purpose of the civil association via the state of nature argument, but he softens this view from its more usual metaphysical implications and seems to read it in light of Hobbes's general "observational" approach (2007, p.31).

That is, Rawls seems to follow J.W.N. Watkins in suggesting that Hobbes aims at producing a kind of political knowledge, which involves our mentally 'breaking down' the state to its constitutive elements in order to see how it works and what it must be for. [43] Hobbes uses this method, Watkins argues, to present a scientifically exact political "science" (1965). While Watkins places a great deal of stress on men as matter in motion to make this work, however, Rawls discounts the necessity of a Hobbesian metaphysics to his political project. Rawls avoids the language of man as an 'engine'— driven on by specific purposes – preferring, as this writing will show, to begin with man as politically driven by fundamental interests. [44]

Rawls is not alone in arguing that Hobbes's core system can be detached from his more "comprehensive" or contestable metaphysics or philosophy. In *Hobbesian Moral and Political Theory* (1987), Gregory Kavka argues along similar lines. With respect to Hobbes's "ontological and metaphysical positions," says Kavka, such conclusions as Hobbes derives are "non-essential" as they can

---

[43] This is the 'resolutive-composite' method J.W.N Watkins (1965) ascribes to Hobbes, and attributes to Hobbes's friendship with Harvey and Galileo. This seems to be the source of Rawls's "reductive method."

[44] Following D. P. Gauthier who uses similar language to Watkins here: man as a deeply "self-maintaining engine" (1969, p.21).

be independently verified (p.11). Although very important to Hobbes's overall view, religious concerns are also said to play "very little role" in what Kavka describes as Hobbes's core "moral and political system" (p.21). Like Rawls, Kavka rejects the idea that Hobbes sees people as essentially egoistic, but for the purposes of Hobbes's political view, he thinks that persons must be assumed to be "predominantly," or usually, self-focussed in their motivations.

Kavka also rejects the idea that Hobbes subscribed to deep psychological egoism, but he did see Hobbes as arguing that we ought to recognize that persons are usually focussed narrowly on themselves, or their immediate families, when it comes to political matters. In these respects, Rawls's earlier lectures will be shown to be in some agreement. Indeed, this view is very comparable to Rawls's "fundamental human interests" reading of Hobbes's political doctrine, but Rawls also has some additional political observations about the need for "universal" reasons in Hobbes's system that he raises but ultimately leaves by the wayside.

In my next section I will expand upon this by giving a view into Rawls's reading of Hobbes's famous "state of nature" argument. Rawls does not view Hobbes as putting forward a view of human beings as being especially negative in their motivations, or even overly egoistic. Rawls does want to show, however, the way that our "fundamental" self-focussed interests animate the state of nature for Hobbes. These reasons are presented as very motivating, and this will later fuel a reading of Hobbesian subjects as "politically" egoistic. This is to say that, on matters of politics they are predominately self-focussed, which creates a particular

problem of stability over how to justify "reasonable" principles of political

organization. This same motivational subset also animates the state of nature.

II

State of nature argument

It is useful to begin by first very briefly touching on Hobbes's basic

argument from *Leviathan*. Recall that, the state of nature argument points toward

our having sufficient reason to want a sovereign authority to establish laws and

security; otherwise, Hobbes tells us, we end up in a scenario of "war." That is, we

live with the knowledge that we are vulnerable to becoming the prey of others,

and this pushes aside all possibility for the fruits of civilized living (XIII). Hobbes

says this would be the case where individuals existed without a sovereign

authority over them because of three sources of potential conflict. These are the

following: diffidence (wariness of the threat of others), competition, and finally

pride (XIII.6). Hobbes thinks the natural wariness we have of these motivations

would only be amplified tremendously if we were aware that no power existed to

reliably secure us from the intentions of others. [45]

---

[45] Hobbes famously asks us to consider our own actions even in relative security as one proof of this. Even in a peaceful society we tend to lock up our things, and take such measures as accord with our own security (XIII.10). This echoes Hobbes's early statement on his method "*nosce teipsum*, read thyself" (I.4). Hobbes allows that one can know the thoughts, passions, and motivations of other men by considering them against one's own on like occasions (real or imagined). One has to be careful though to take specific account of circumstance, so that one does not judge with too much "trust" or "diffidence" (ibid). Hobbes states that his full account works like this; that is, it offers up a view, ordered and well considered, that other men must ultimately try to read in themselves and judge. This would seem to resonate with Rawls's reading of Hobbes (and Kavka's) that allows them independent persuasive force apart from metaphysical or religious interpretation.

In authorizing a sovereign what we do is renounce our absolute right of nature to all means and actions aimed at self-preservation and also our position as absolute judge in all things with respect to this. This "second" law of nature flows naturally out of the primary one for Hobbes, which is that we ought to strive to preserve our own existence (XIV.3). Allowing the sovereign to restrict our liberty and to make judgments about right conduct, and to make laws that all must obey, is the primary means to effective self-preservation and also the only means to all the fruits of social living. The opposite condition would be a war between individuals who must use whatever means they can try to best preserve themselves, and their limited family lives, against others who are likewise absolute judges in their own case and at absolute liberty to all means of self-preservation.

Rawls's interpretation stresses this last point particularly. According to Rawls's lectures, pride and competition should not be seen as straightforwardly sources of conflict in Hobbes's state of nature. Indeed, Rawls thinks not many are likely to be moved to conflict by these motives, or at least we do not have to assume they would be (2007). (In terms of competition there is scarcity, but Rawls tells us that we do not have to assume that a reckless drive for "power after power" necessarily turns scarcity into conflict [p.50-1]). However, Rawls points to the mere possibility that these intentions might hold with at least some people giving us reason enough to be afraid of others where there is no law to restrain them. Hobbesian subjects thus become wary of each other not just as rivals but potentially fearsome and deadly rivals. For Hobbes, again, we are both our own

judges of what is necessary for self-preservation and entitled to all means of self-preservation. Rawls makes the fairly intuitive judgement here that individuals may seek to use the means of conflict to eliminate the threat of their rivals, and (worse) the supposition that others may be thinking the same thing only encourages such destructive diffidence further. [46] (Hence we have the idea that the state of nature is essentially a tragedy).

Rational actors acting against the backdrop of uncertainty causes conflict in this scenario. Choosing a sovereign is, under such circumstances, what Rawls terms "collectively rational"; that is, given our desire for peace and civilized living, it is rational for us to forgo unlimited individual judgement as to the means of our self-preservation in favor of a common authority if others do likewise. The alternative is a conflict that is driven largely by fear (2007, p.48-58). Thus, while certain impulses might tempt an individual to defect from an agreement to have a sovereign, our 'fundamental human interests' can never be reliably protected this way. Where others are willing, we should renounce our claim to all things and agree to let a sovereign impose and enforce laws.

Hobbes is then said to offer a further set of propositions that are presented as promoting civil peace. These are Hobbes's other laws of nature and are viewed by Rawls, along with Hobbes's first laws of nature, as reasonable in their content. Although Rawls does not go deep into these, they are summed up by the Hobbes's maxim that we should not to do onto others as we would not have done onto

---

[46] I use a bit of artful language here for Rawls, but I believe this articulation to be in line with his section "The argument for Hobbes's Thesis" (2007).

ourselves (XV.35). That is, besides recognizing equality and avoiding *pride*, one

also keeps one's agreements (which is *justice*); one accommodates others by not

taking more than a fair share – especially when others are desperate; one tries to

be modest and does not act out of arrogance or a need for revenge; [47] one deals

with equity between others; one pardons when possible, and one follows

Hobbes's other terms of peaceable association (XV.1-41). The grounds of these

however is ultimately a need to protect our essential interests, which also amplify

the trouble that the state of nature represents.

Indeed, for Rawls the crux of the state of nature argument for collective

action is this: "we do not have to be monsters to be in deep trouble" (2007, p.51).

Even "decent" people might destroy each other absent some force to keep

everyone in awe (ibid). Subjects want to protect their lives and those things –

family life and means to prosperous living – that are essential. But in a state of

nature we have a disastrous collective outcome due to something similar to what

A.P. Martinich (2005) elsewhere calls "The Great Fear and Danger Argument," [48]

which goes something like this: (1) some people are dangerous in the state of

nature. (2) It is not possible to know which people are dangerous. Therefore, (3)

it is necessary for a person to be afraid of everyone in a state of nature (p.70). The

---

[47] On the point of accommodation, Hobbes speaks of those who acquire for themselves what is "superfluous" to them but "necessary" for others (XV.17). This is interesting in light of the fact that Hobbes is sometimes pointed to as prefiguring the modern "capitalist" world. C.B. MacPherson for instance memorably claimed that the Hobbesian subject of Leviathan was essentially Bourgeois, and that Hobbes placed no moral barriers upon his material acquisitive drive (1965). In contrast, Keith Thomas (1965) points to the accommodation clause in *Leviathan*, and others, to advance a more nuanced view. Indeed, in order to maintain civil peace, some level of subsistence for all men may be justified (and further no one can claim a right of private property against such, as no such right can be held against the sovereign).

[48] Martinich thinks this is at work in *Leviathan*'s "state of nature" argument and observes that the similar view presented in *De Cive* makes this clearer.

summation of this argument is that "universal fear does not require universal danger" (2005, p.72). Thus we are spurred on by destructive diffidence. [49]

The state thus comes into existence to solve this kind of a "collective action" problem. Rawls maintains a focus on this, suggesting further that Hobbes sees a "prudential" securing of our most essential interests against "advantage seeking" destructive behavior – threatening to return us to a state of nature – as being Hobbes's main justification for the political compact (2007, p.34). That is, persons realize that there is a collectively rational reason to follow the laws of nature (especially obeying the sovereign), and they can feel secure that observing those actions is what they ought to do because there is a sovereign to ensure everyone has reason to comply with the generally socially constructive scheme.

Rawls is careful in his appraisal of Hobbesian human nature in this. He is deliberate in not suggesting that Hobbes thinks that humans are psychologically "egoistic" as a rule generally. [50] What Rawls does end up saying, however, is that "politically" the influences of human behavior attach to fundamental self-focussed interests enough that persons seem to be predominantly egoistic on matters of

---

[49] Thus Rawls tells us that Hobbes does not need to base his theory around "pride" or "vainglory" or other negative motivations with respect to the state of nature. According to Rawls, "we can say that if pride and vainglory and the will to domination is a possibility, then that is strong enough for our purposes … the difficulty in the state of nature is the great *uncertainty* about the aims and intentions of others. As long, then, as love of domination and vainglory are *psychologically possible,* these passions are a complicating factor in the state of nature. A general state of uncertainty about other's aims and intentions characterizes the state of nature, so that a concern for our self-preservation forces us to consider the *worst* possibilities" (2007, p.50).

[50] Following a line similar to Kavka's definition, this would mean that subjects always intend through their actions some objective good that is self-focussed. But this is not the same as saying that everything an individual does they do because they have some desire or "motivation" to do so originating from the 'self '(1987, p.35). This last is what Kavka calls "tautological egoism," and he takes it simply to be a truism (ibid). Kavka suggests Hobbes fell somewhat unwittingly into a "subtly" similar category to the latter "tautological" view, but he did not intend that subjects are egoists in the first more strongly 'self-centred' sense.

politics. There are two reasons for this reading. The first and most dominantly

stated is a motivational account suggesting these are the most compelling sorts of

'Hobbesian' reasons, and the second is expressed through Rawls's frequent use of

language that suggests that these are the only sufficiently *universal* reasons as to

be politically useful for Hobbes. To begin to get at this division, my next section

will look at how Rawls treats Hobbesian human nature.

<div align="center">III</div>

<div align="center">**Hobbesian human nature**</div>

Rawls interprets *Leviathan* as putting forward a prudentially rational view of

political life. This does not mean however that Rawls takes Hobbes to be positing

that persons are fundamentally egoistic. Instead, Rawls wants to be very careful to

show Hobbesian subjects as having a diverse range of possible motivations,

including those complex moral motivations one might garner from living in

society. What is central to Rawls's interpretation, however, is that the Hobbesian

perspective on political life takes subjects as if they *were* egoistic. This is because

the most important and universal human motivations, those which explain the

state, are largely self-focused and not social. This "self-focus" also explained for

us why decent persons might nevertheless come into conflict without a sovereign

to guarantee their cooperation.

Nevertheless, Rawls wants to begin by rejecting the view that human beings

are innately bad creatures that some might glean from certain sections of

Leviathan. Rawls does not take Hobbesian subjects, strictly speaking, to be

egoistic. [51] This is stated outright, and Rawls also goes further in observing that

Hobbes allows for the possibility of "benevolence" and a love of justice for its

own sake in his account of human nature as well as "good will" and "charity"

(2007, p.45-6). For good will and charity a number of places are referenced in

Leviathan, particularly chapters six and thirty (for a particular emphasis on

conjugal affections). Rawls also cites chapter fifteen as saying a person might

have a certain relish of justice, which is a courageous disposition that scorns to

have the contentment of one's life depend on "fraud, or breach of promise" (2007,

p.46). If Hobbes is one of Rawls's "dark minds" in Western thought, then, it is

not on a view of human beings as necessarily 'wicked' in nature. Indeed, a

Hobbesian view of human beings, for Rawls, includes much to be approved.

What is more, Rawls sees a Hobbesian view as allowing that people are

capable of having an array of socially driven motivations; "social institutions and

education and culture" can all change human passions in significant ways in

Hobbes's system (2007, p.42). Rawls soon repeats this observation about the

social nature of motivations working on human subjects, altering the course of

their passions and desires, which he notes Hobbes would not have failed to notice

---

[51]These are a few of Hobbes's more famous articulations that men have a "restless desire of power after power that ceaseth only in death" as well as the idea that men are bound to "invade and destroy one another" in a state of nature (XI.2, XIII.10). Hobbes does indeed check these observations himself. All persons may desire power, but they do so only because one can never be secure enough in what one has. Pure love of power as *domination* is only a desire to be found in "some" but not all, and the power to acquire what we want can be judged against the objects of our desires themselves. Successful rulers, for example, may desire fame from conquest, but others may prefer ease, or flattery for some domestic achievements such as in the arts, or any number of other possible motivations (XI.2). With respect to the state of nature the 'good news' of course is that our passions, including our desire for commodious living, also incline us to adopt such "articles of peace" as "reason" suggests (XII.14).

(2007). Here Rawls is being careful in allowing Hobbes a nuanced and thorough approach. Although a certain deficit in the efficaciousness of such sources of motivation is important Rawls's interpretation, Rawls does not want to suggest that Hobbes offers a less than credible presentation of human nature that would exclude its obvious social and moral aspects.

Rawls does however temper down his generosity in explaining why such sources of motivation are not politically decisive for Hobbes's thought. For the purposes of Hobbes's political doctrine, or "secular moral system," what counts are the *essential* features of human nature, and Rawls suggests that these are largely self-focussed. These are the basics of self-preservation, the means to one's own contented life, and the securing of one's conjugal affections. Further, Rawls later goes on to argue that these are the "most important final ends" we have according to Hobbes (2007, p.56-7). They are largely self-focussed, and they are not socially derived or "principle dependent." They are in fact interests we would have even in a state of nature.

At the same time, Rawls also emphasizes Hobbes as downplaying other sources of motivation and aspects of human nature. The love of justice for itself is, for example, something that is "rare," and the psychological makeup of human beings is still predominantly "self-centered" when it comes to "basic political and social matters," where human beings aim primarily at securing their essential goods (2007, p.45-6). Here Rawls pulls back to suggest that, although Hobbes allows much that Hobbesian subjects need to have in terms of the kinds of motivations required to adopt a moral view of the political compact, these are

simply overwhelmed by other kinds of (more self-focussed) considerations. Indeed, for the purposes of Hobbes's political doctrine, or "secular moral system," what counts are the essential human interests.

Hobbes's simplified political take on human nature is here presented in order to best give an account of what "holds civil society together" (2007, p.46). Although one might find a range of possible motivations in human nature (especially as affected by civil society), these are typically not strong enough to rely on as a basis for social unity. In order to achieve social unity, we must introduce a class of interests that are common to everyone. The fundamental human interests are "final ends … states of affairs or activities which [people] strive for their own sake" (2007, p.51-7). Rawls is rather loose here in alternatively stressing that Hobbes builds his thought around these because they are "most important" or rather because they are "common." Both adjectives are used, but Rawls does not want to tell use which is more important for Hobbes's political view. [52]

Toward the language of the "universal," Hobbes's political doctrine does seem to be taken by Rawls as mobilizing 'agent neutral' reasons as the basis for political unity. They would be applicable to any acting agent, regardless of any other information. On Rawls's view, Hobbesian subjects may have many possible

---

[52] What Rawls is keen to say is that these desires are object dependent in that they can be "described without mentioning any reasonable or rational principle, or moral notions generally" (ibid). There is, for Rawls, no reference to a desire for being treated justly or other notions that might be moral, or "principle dependent," desires (2007, p.58).

motivations gained from society but these do not, of course, converge (2007). Hobbes simplifies then by tying the political life to reasons that *do* converge, and these revolve around our final ends as well as principle dependent desires that are derived from these; that is, a desire to act rationally in accordance with what serves our fundamental interests (the best means to our shared ends). Principle dependent desires that are desirable because of being reasonable, or "moral," are excluded on this view, as they are not derived from our fundamental interests (2007, p.58).

On Rawls's interpretation, Hobbes places an emphasis on the view of society 'as if dissolved' in order to get a better view of such interests as seem to be universally compelling. In looking at the state as if dissolved Hobbes is able to get a view of human nature that stresses our 'fundamental human interests.' It is action from the motivations to protect such interests that shows both why the state of nature would be destructive, and what the state might have been 'created' to do (secure these interests). Rawls's interpretation here is that Hobbes presents a view of "rational prudence" that is political from this in its emphasizing these core 'self-focussed' concerns. Before moving on, I will offer a little bit more on the "political" foundations for rational prudence.

IV

Rational prudence and "predominant egoism"

Rawls's view of Hobbesian subjects as 'prudentially rational' owes something to D.P. Gauthier (1969) in his reading of *Leviathan*'s laws of nature.

Rawls notes this in passing in his lectures, noting also his intention (arguably unfulfilled) to sidestep the well-known controversy in Hobbesian scholarship over whether obligation in Hobbes's system might indeed have a moral dimension (as for example Warrender and Taylor argue [2007, p.37n]). Rawls does not say a lot more about his use of Gauthier, but it is very clear that Rawls follows him in saying that there is no moral basis for obligation in Hobbes's system. Indeed, for Gauthier such obligations as Hobbes's laws of nature may establish for us, cannot stand independently of Hobbes's psychology. It is this that leads to Hobbes's stressing his principles in terms of serving individual advantage. (Gauthier reads Hobbes's psychology as ultimately destructive of *Leviathan*'s ethics).

For Gauthier, the psychological dictum that each individual always chooses for his own advantage (which he sees Hobbes as endorsing) makes it impossible for the laws of nature to bind as ethical obligations. On this view, "the conditions under which man has an obligation are not those which we classify as moral, rather [they are] merely prudential" (1969, p.97). Indeed, the use of 'prudential' there should stand out, as Rawls is going to end up sounding a lot like Gauthier. According to this view in Hobbes's system, "both rights and obligations must have a prudential foundation," and they must always be "instrumental to the ends of the individual, which are self-centered ends" (1969, p.93).

This falls short of moral obligation for Gauthier, as obligation entails that one ought to do something whether or not it is to one's advantage. Of course the catch might be that Hobbes suggests that what one ought to do and what is to one's advantage are necessarily always in sync, although Gauthier doubts that this is so.

[53] Rawls follows Gauthier in suggesting that Hobbes's system is one of 'rational prudence.' But Rawls softens Gauthier's view slightly by the suggestion that, as we have seen, that Hobbes's focus on "self-centered ends" is mainly for the purpose of his *political* conception. Hobbes does not argue that reasonable motives or those based on mutuality exist, but on this view he has chosen to focus on interests he thinks all persons might share, or find compelling, in order to achieve social unity. Other motivations may not only fail to be likewise universal, but may also simply not be strong enough to give a credible account of how political life hangs together according to Rawls's reading (2007).

Rawls's account rejects a strong psychological account of Hobbesian subjects as driven primarily by a kind of self-interest. He does not use Hobbes's metaphysics of matter-in-motion to underpin his explanation of why the fundamental interests are important to us. Gauthier, on the other hand, founds his argument on Hobbes's psychological account of man, which ultimately relies on Hobbesian metaphysics of matter in motion and the idea of man as a self-perpetuating engine whose primary "design" is to preserve himself. [54] As I related at the outset, Rawls wants to reject this grounding for the purposes of Hobbes's

---

[53] Gauthier (1969) introduces some examples to suggest there may be a number of cases where what is most reasonable is not to the direct advantage of all concerned. Gauthier concludes that "what is reasonable and what is advantageous cannot be completely coincident" (p.96).

[54] According to Gauthier (1969), "Hobbes undoubtedly did not recognize the difference between physics and politics, between the connection of gears, wheels, and levers with a mechanical engine, and the connection of men with a Commonwealth … if we accept Hobbes's view that man is a self-maintaining engine, then we can establish the nature of human motivation. Men want, and necessarily want, to preserve themselves. Therefore, whatever can be shown to be a condition of human preservation, is thereby shown to be a means to man's end" (p.21).

political doctrine. Rawls thus reads Hobbes in a similar manner to Gauthier but does not follow him all the way down in his assumptions.

Rawls again, though, is not entirely clear about his own political "egoism" assessment, and why thinks that this is the basis of a Hobbesian "political doctrine." His statement that Hobbes is trying to "model" the major psychological forces in politics seems backed up by his assertions that, again, these are our most compelling ends and that the state comes into existence to protect these ends, which also cause the state of nature to be so destructive. Lost in this assessment are articulations over these being "common" interests, which fit into a view of Hobbes trying to address another kind of political problem – that pluralistic subjects need some grounds on which they can agree to form the political compact.

Instead Rawls leans on the idea that Hobbesian subjects are predominantly egoistic on political matters. Hobbes thus needs to emphasize these aspects of human nature, though he acknowledges that these do not provide an exhaustive account of human motivation. Rawls goes on to say what role the sovereign plays in this situation using terms very familiar to his own work. The sovereign "stabilizes" social cooperation by providing "public knowledge" that everyone in society has reason to follow the laws of nature (2007, p.73). Hobbes therefore solves a similar 'assurance problem' familiar to Rawls's own thought. According to this idea, everyone can be given a view of what is to their "mutual advantage," parsed in terms of their most fundamental interests, but they need further assurance that it is rational to act accordingly. Thus public knowledge that there is

an "*effective* sovereign" gives everyone the sufficient reason they require to comply (2007, p.79). [55]

Although Rawls introduces this idea in (what is arranged as) lecture IV, readers do not get substantive comment on this concept from Rawls's own perspective. Comment on Hobbes's view of the political compact, from Rawls's perspective, is only substantively offered in lecture III, which addresses Hobbes's "practical reasoning." This is largely a "moral" critique of Hobbes's approach. It follows a Gauthier like line in objecting to Hobbes's lack of reliance on "moral obligation," and it leaves a more complete political reading of Hobbes's thought behind. The significance of this last I will raise fully in my next chapter, but I will conclude here by outlining the critique Rawls does offer.

V

Rawls's moral critique

Rawls raises the distinction between "rationality" and "reasonability" again in assessing the Hobbes's view of the political compact. Rationality here "involves furthering the good or advantage of oneself, or of each person cooperating" where such a disposition is distinct from what Rawls terms "the reasonable," which tends to imply things like "being fair-minded, judicious, and able to see other points of view" (2007, p.54). The distinction is illustrated particularly in Rawls's lectures through the example of someone who drives a

---

[55] Rawls presents a game theoretic explanation here in terms of 'payoff tables' that may remind one of *TJ* (2007, pp.78-9).

hard and "unreasonable" bargain. Such actions are rational from a narrow view of one's concern for one's own individual good, but they are obviously not the same as that which springs from a person's desire to act from a "principle dependent" set of fair terms of association (2007, p.54).

Rawls's view of Hobbes's political doctrine is 'prudentially egoistic,' as has been related. One ought to have a conservative approach in following the laws of nature, and 'advantage seeking' in ways that are detrimental to these laws is potentially radically destructive of our fundamental interests. What Rawls does with this view of things is primarily evaluate it from the perspective of moral philosophy, drifting into a Kantian assessment as he goes. Hobbes errs, according to Rawls, by presenting terms of social organization that are not "categorical" in nature. That is, they are not affirmed on their own grounds (2007). Hobbes's system instead relies on *assertoric hypothetical* imperatives, which are based on secondary considerations; that is, "they are valid for all on the grounds that we all have, as rational beings, a certain end" (p.64-5).

Stated more plainly, Hobbes's system – whatever the reasonability of its content – is based on secondary considerations of self-preservation and the "means to a contented life" rather than a primary desire to act on reasonable principles. Rawls's point is that Hobbes has based his "secular moral system" from a system of practical reasoning where ostensibly moral conclusions rely on purely rational imperatives (hence the reasonable is derived from the rational), and Rawls objects from the position of a theorist who wants to mark out

independent terrain of practical reasoning for reasonable principles themselves,

something he thinks Hobbes does not want to do. [56]

Hobbes's "practical reasoning" gave that political life, while excluding a

union over a common good, proceeds on the basis of avoiding society's *summum*

malum (conflict). On this view, Hobbes provides a *rational* basis for the

organization of the social world, where the outcome of the laws of nature is

"collectively rational" rather than reasonable; that is, again, when everyone

follows them and this observance is publicly known everyone has reason to do

likewise, given their interests (2007, p.64). This view empties Hobbes's laws of

nature of 'moral content' for Rawls.

While the principles of justice that would structure a well ordered society

may be 'free standing' from any one comprehensive view, they nevertheless must

command categorically, or else we cannot be said to be truly cooperating. This is a

general concern and phrasing that appears in Rawls's lectures, although it does

also show up (in somewhat weaker form) in *PL* (2005). Instead, according to this

view of things, we are merely "coordinating" according to some hypothetical end.

The distinction here is between acts that are in *conformity* with duty and acts that

are specifically done *from* a sense of one's duty, a roughly Kantian concern.[57]

---

[56] According to Freeman (2007), this is indeed one key difference between Rawls's social contract and "Hobbesian positions." "Rawls's social contract differs from Hobbesian positions primarily in its claim that reasonable moral principles cannot be derived simply from the concept of rationality, or a person's good. Rather, the Reasonable forms a distinct and independent domain of practical reasoning with its own independent moral principles" (p.345).

[57] By "categorical" is meant acted upon principle because we recognize our moral obligation. This echoes Kant's "test" for moral validity. Kant rejects that the proper purpose of "reason" would be for self-preservation or happiness, as "instinct" is arguably a better instrument for such purposes (2005, p.56-7). Kant argues instead that the purpose of reason should be to produce a will that is *good in itself,* rather than as a means to an end (ibid). To authenticate whether one has acted with a

This is also the difference between social cooperation and what Rawls terms political life organized on the lines of social *coordination*. Social activity that occurs in the latter is rule bound and based on advantage, and Rawls thinks it akin to "bees in a hive," or, more usefully, to "workers" in a factory (2007, p.56). This is distinct from a society that may be organized along the lines of "social cooperation," which proceeds for the advantage of all but is first limited by *reasonable* terms of association (with associations of reciprocity and mutuality) that are endorsed for their own sake as being fair and that may require some to forgo their immediate advantage.

Rawls thinks Hobbes's conception leaves no room for the idea that political subjects might be moved from "a sense of fairness." Instead Rawls sees Hobbes's "secular moral system" as attempting to base itself entirely off the 'moral power' of "rationality," or our ability to figure out what is to our advantage. Thus, the idea that cooperating might be good for its own sake is excluded. This matches the idea of a '*modus vivendi.*' According to this view, society is like a compact between separate parties for mutual advantage, which is how Rawls describes Hobbes's system. They endorse the terms of their association because they serve their self-interest, but they would surely choose some other arrangement if it suited their interests better. The arrangement is chosen because it works, but it

---

good will one must check to see if he or she has recognized and acted from his or her moral duty. Our actions must come from our obligation to do our moral duty rather than from some other inclination. Here Kant gives the memorable example of the tradesman who does not overcharge, but keeps the same price for everyone (even an unsuspecting child). If the tradesman did this from his sense of moral duty, his actions would have moral content, but if some other motive drove the tradesman (a fear for his reputation or even love for his customers perhaps) his actions would not have moral content (2005, p.59). One's moral duty in this case is determined by Kant's familiar 'categorical imperative' procedure.

does not win out through a sense of moral allegiance as that would suppose that the terms of their association were endorsed in principle as fair or otherwise worthy.[58]

Hobbes's view of political life excludes the possibility of a moral political consensus, then. That is, it makes too little room for the idea of citizens acting socially according to mutually recognized fair terms of cooperation (as legally structured) that are recognized as reasonable and affirmed on that basis. The idea of 'social cooperation' that appears in Rawls's lectures is not exactly an 'overlapping consensus,' as Rawls does not quite show the awareness of the need to adopt such a concept. His critique of Hobbes merely contrasts a "moral" versus a "non-moral" vision of the political compact. It raises but does not fully address the contrast in the problems of consensus, stability, and assurance, as questions of political philosophy.

VI

Chapter assessment

Rawls can be seen as raising a number of terms and ideas familiar to his later work in these lectures.   Hobbes is said to have a "political doctrine" that

---

[58] The analogy to workers in a factory that Rawls makes is exemplary. Workers show up at a factory because it serves their interests to do so. The terms of their employment are accepted merely because they work for them now, even if the reasons might not be very good, such as in the case of economic dependence. The terms might be abandoned if something better (like a new job) came along. Society, according to this view, is like a compact or treaty between different parties on a mode or way of living that is accepted because it serves the interests of all, for the moment, but that is all. (The case of the workers is 'social coordination' as *modus vivendi,* while the situation with the bees is arguably something different entirely.)

emphasizes certain aspects of human nature. Hobbes wants to tie his subjects

together around the "laws of nature" in this regard. These suggest primarily the

role an effective sovereign is to play in structuring society. But Hobbes's other

laws of nature are important as well, and they suggest reasonable terms of

reciprocity. These however need to be enforced by the sovereign. The sovereign

solves a kind of problem of stability by creating public knowledge that these laws

are enforced, and thus everyone has reason to follow them. That is subjects have

reason to want these to be enforced because it can be known that such a situation

serves the mutual advantage of all, springing from a motivational focus on our

common and fundamental interests in self-preservation, protecting our family

lives, and securing the means to live well.

In this Rawls assumes two things in particular. (a) A sovereign is required to

stabilize social cooperation rather than a public view of justice that is endorsed by

everyone, subject to some limits. [59] And (b) the ultimate justification for such a

social compact attaches to "political" reasons Hobbes thinks all must accept.

Rawls does not rely on a metaphysical conception of Hobbesian subjects, or a

deeply psychological account, that would make them egoists for this. However, he

does lean heavily 'predominant' egoism or a view of politics that needs to take

into account the most significant features of human behavior. Rawls suggests two

reasons Hobbes leans on this approach. The first is that these are our "most

---

[59] Of course, some coercive enforcement is still required for things like the payment of taxes. But the primary assurance that the "public" conception of justice is enforced is the collective loyalty of citizens to it, stabilized by a public perception that this is the case, as opposed to the knowledge that a sovereign exists to enforce punishments for non-compliance.

significant final ends," and thus are highly motivating; the second, less

prominently stated, is that these are the most common, universal, and thus perhaps

only basis for some kind of a "freestanding" political view (if not a moral one).

This last approach is not systematically worked out by Rawls. It is mainly

suggested by many of the adjectives Rawls uses in evaluating Hobbes's doctrine.

Rawls does however find many places in his lectures to note where Hobbes is

troubled by the political dimension of value pluralism in society (touched on in

my next chapter), but he does not work these into the account of the stability of

the Hobbesian political compact. In my next chapter I suggest that no view of

Rawlsian and Hobbesian political consensus and stability is complete without

taking this dimension into account.

Chapter three: Rawls and Hobbes and the question of pluralism.

Introduction

Rawls's assessment of Hobbes's *Leviathan* had attributed some "political features" to Hobbes's vision of the social compact. In Rawls's view, Hobbes's core doctrine is intended to stand apart from any "comprehensive" claims but remain persuasive as a set of reasonable principles of reciprocity. Hobbes leans on "political" prudential-egoism for his primary justification in this, as self-focussed concerns are taken to be important enough that Hobbes must use them predominantly to "model" human political behavior. This fueled what I called a "primary problem" of stability familiar as well to Rawls's early work – that being, how can we be assured that others will do their part in maintaining a system of political reciprocity and that they will not defect out of advantage seeking or direct self-interest? Phrased another way, *Leviathan* asks how political subjects can be relied upon to do what is mutually advantageous, even if they might worry that others cannot be relied upon to not act in "selfish" (ultimately self-defeating) ways.

Rawls then goes on to critique this approach on grounds familiar to his moral philosophy. Lost in this assessment, however, was what importance one could attach to the fact that Hobbes was said to stress 'universal' or 'common' reasons with his focus on "fundamental human interests." In this section I argue that Hobbes's "political" focus in that regard, which Rawls raises but does not

work out at length, is informed by the requirement to solve a second problem of stability familiar to Rawls's later work. That is, how can we have a stable consensus amongst political subjects who are also deeply irreconcilably pluralistic? I argue that it is this pluralism that ultimately makes certain "social" learning and motivations deeply problematic for Hobbes, and explains more fully his "political" focus on "predominantly egoistic" interests that Rawls points toward (although Hobbesian subjects certainly need not be seen as egoistic in a deep psychological or metaphysical sense).

Rather than explore this view, Rawls leans on the idea that Hobbes must have thought human beings were predominantly egoistic for the most part, even if not in a thoroughgoing sense. Although Rawls is quite aware of how problematic societal pluralism is for Hobbes (as we will see), his critique misses the contribution of pluralism to Hobbes's political conclusions. Indeed, complex social, political, or otherwise comprehensive, motives are quite significant for Hobbes's view, but their impact is largely adverse. That is, they create a significant problem for the stability of political life, threatening to create faction and even social strife. It is this that drives Hobbes's "political" focus on universally held interests, which is not the same as saying that *Leviathan's* doctrine assumes that persons are predominantly egoistic without further comment.

If predominant egoism held simply, it would just be a matter for an effective sovereign to prevent "advantage seeking" defection from cooperation and society would be stable. But there is a second stability problem at work in Hobbes's

thought. What if, out of loyalty to rival conceptions and allegiances rooted in religious, political, or philosophical views, subjects were motivated to defect from social cooperation? Such views could lead to defections either (a) directly through what Hobbes terms "pride" or great "vainglory," or by (b) creating a sense that others challenge the state and do not recognize the sovereign, making others less likely to want to "do their part" in terms of social cooperation.

Ultimately the problem Hobbes sees here is so deep that *Leviathan* does not finally give us reason to think that it *can* be solved in a political way even if Hobbes's 'core' recommendations do have a political shape. That is, if Hobbes does try to present a "free-standing" view, he nevertheless clearly retreats from suggesting that such an approach could stand in as a basis for political consensus in a way that would match a Rawlsian model. Where Rawls is willing to cast pluralism as "reasonable" and to hope for an overlapping consensus, Hobbes in his time did not think the problem of stability raised by pluralism could be solved except by authorizing an absolute sovereign to represent everyone.

Indeed, Hobbes's skepticism over the kinds of interests that could unite people politically also informs an apparent pessimism that his own laws of nature – however grounded – could serve as a free standing political doctrine. In failing to recognize this element of Hobbes's thought, Rawls misses an opportunity, early in the development of his own political liberalism, to drive towards the 'second' problem of stability that arises with the assumption that society is irreducibly pluralistic. Rawls thus misses an opportunity to juxtapose and defend his own

conceptions of "reasonable" pluralism and an overlapping consensus with Hobbes's earlier more 'conservative' vision.

The difference does not turn on a question of individual moral natures as simply a matter of "egoistic" versus "reasonable" temperaments. Rather it asks an arguably more challenging question over whether a political and "moral" vision can be shared by subjects who might have a diverse range of motivations, but who are irreconcilably pluralistic in their comprehensive outlooks. Rawls's lectures, then, do not contain a reply to Hobbes's work on the level of its most significant challenge to Rawls's later political philosophy. Is *reasonable* pluralism a reality, and might we have faith that an overlapping consensus is possible? This is a point that I will return to by this work's end. I will start, however, by clarifying a few important terms and concepts that have been raised so far.

I

Some preliminary clarifications

What is meant by "political" throughout has been fairly broad in scope. It is intended to mean some basis for people who do not share a common view of the "good," or "comprehensive doctrine," to be able to come to a consensus on the terms by which they organize and govern themselves. This covers Rawls's moral approach, but *also* what Rawls terms as "political in the wrong way" (2003, p.189). A doctrine is "political in the wrong way" when it looks for some overlap or coincidence of shared interests between comprehensive views as being the basis for political agreement (ibid). Rawls has elsewhere criticized this as the

basis for what he calls "Hobbesian liberalisms," which try to balance interests and

advantages to reach an agreeable arrangement. [60] This work roughly takes Rawls

as intending that Hobbes's core view has "free standing" aspects in this latter

sense. [61]

Rawls's own view of a "free standing" conception has a moral dimension.

Justice as fairness is an example of a liberal political conception that is morally

affirmed on its own grounds, and does not depend on any one particular

"comprehensive doctrine" as its source. "Free standing," for the purposes of this

work is broadened simply to the idea of a view that is based around terms that

stand apart from comprehensive doctrines. While Hobbes's core "laws of nature"

are given a philosophical and religious defense of sorts, it is clear that Rawls

thinks that these could stand apart as reasonable propositions on their own.

Ultimately, Rawls criticizes the means of justifying such principles, but he ends

up on less than helpful terrain. With his moral critique, Rawls comes closest to

saying Hobbes's system merely fails a "test" of moral validity on specifically

Kantian grounds.

The moral status of Hobbes's "doctrine" is not debated here. Rather, this

work wants to ask what "political" questions Rawls has passed over in assessing

Hobbes's system. It is argued here that Rawls's take on Hobbes's "political"

---

[60] Rawls is illuminating here. By the "Hobbesian strand of liberalism" he means the idea that "ordered liberty is best achieved by skillful constitutional design framed to guide self-(family-) and group interests to work for social purposes by the use of various devices such as balance of powers and the like … This strand becomes purely Hobbesian to the extent that it sees self-(family-) and group-interests as the only available, or the only politically relevant, kind of motivation." This is taken from "The Idea of an Overlapping Consensus*"* (1987, p.422).
[61] By 'core view' I again mean Hobbes's state of nature argument and program of the "laws of nature," following Rawls's lectures.

egoism must rely on an assumption that persons are normally irreducibly pluralistic, and that a justification of the state needs to begin from premises that everyone can be relied on toaccept. Hobbes's view is epistemic in this respect, or knowledge ("science") based; it aims to clarify certain principles starting from the premise of self-preservation and equality in a state of nature. Hobbes is focusing on what political subjects share because their other sort of more thickly "social" interests are not likely to converge. [62]

Thus Hobbes begins with a core approach that focuses on what "all men can agree upon" – that peace is good and that the ways of attaining civil peace are knowable and desirable principles of social organization (XV.40). This method is driven by a 'second' problem of stability in Hobbes's work, which Rawls points toward, over how to justify (whether it is at all possible) the political compact to persons who can differ so comprehensively in their deeply held moral and philosophical views. Hobbes must deal with human beings as potentially "self-interested" and "advantage seeking," then, but also as complicatedly social and difficult to reconcile on "comprehensive" grounds. [63]

Rawls's later article "The Idea of an Overlapping Consensus" (1987) captures Rawls more clearly reading Hobbes in this latter respect. It offers a

---

[62]This work passes over the question of whether Hobbes intends to refute moral "skepticism" with this gesture in a similar vein to Grotius, as Tuck (1988) argues. A discussion on this point can be found in A.P. Martinich's *Hobbes* (2005, pp.229-32).

[63] The basic line of argument is accepted, but it is noted that certain features of *Leviathan* cannot be ignored. Nearly half the work is a work of theological interpretation, which commentators such as S.A. Lloyd (1992) persuasively argue is integral to Hobbes's project. Hobbes very arguably did not think his work could stand without a religious defense, but this work follows Rawls in assuming that Hobbes's core doctrine is essentially "political" at least in potential.

sharper focus on pluralism as a political quandary in Hobbes's thought and the

places "self-interest" in that context:

> When Hobbes addressed the contentious divisions of his day between religious
> sects and between the Crown, aristocracy, and the middle classes, the basis of his
> appeal was self-interest: men's fear of death and their desire for the means of a
> commodious life. On this basis he sought to justify obedience to an existing
> effective (even if need be absolute) sovereign. Hobbes did not think this form of
> psychological egoism was true; but he thought it was accurate enough for his
> purposes. The assumption was a political one, adopted to give his views practical
> effect. In a society fragmented by sectarian divisions and warring interests, he
> saw no other common foothold for political argument (p.422).

Here we see Rawls articulate a view a lot more in line with the "political"

requirement to find some practical basis on which pluralistic and divided citizens

might agree. This is stated plainly by Rawls, and there is less of a reliance on an

assumption of predominant egoism as simply 'true' in an unqualified sense. This

is not to say that Rawls would abandon the idea that Hobbes thought persons were

very often egoistic in motivation, but it is possible to appreciate most clearly here

the contribution of the idea of social division to Rawls's reading of Hobbes's core

doctrine. [64]

In this work's next chapter, I take up the question of "social interests" as a

political problem in *Leviathan*. Hobbes shows an emphasis on the "second"

problem of stability where he raises pluralism in terms of the problems it raises

for the political association. Hobbes must ultimately emphasize those points on

which men can agree, because there is otherwise such a broad terrain on which

they do not find such consensus. Indeed, if anything the tendency of people to

---

[64] Interestingly the reworked version of this article that appears in *PL* excises the references to
Hobbes (2005).

disagree in society over matters of politics, religion, and other concerns, is greatly

troubling for Hobbes.

<div align="center">II</div>

<div align="center">

The problem of pluralism in *Leviathan*

</div>

It is worth noting that a number of thinkers have indeed pointed to the

similarity between Rawls's later work and Hobbes's own on the point of

irreducible pluralism. John Gray for one raises this similarity in his *The Two*

Faces of Liberalism, praising both Rawls and Hobbes in this respect (2000). Gray

in fact contrasts the two thinkers in a manner familiar to Rawls. Where Hobbes

tries to balance forces using mutual self-interest (think Hobbesian liberalism),

Rawls relies on the historical emergence of certain democratic values to suggest a

moral vision of politics beyond a *modus vivendi*. Gray prefers a solution derived

from Hobbes's premises, which assumes that some balancing of conflicting views

of the good must take place, rather than assuming they could be united by an

overlapping consensus in a Rawlsian fashion. [65]

---

[65] For Gray (2000) modern political subjects are simply too divided on questions of what are authentically 'human interests' to ever be reconciled. Even a value like freedom, for Gray, must be weighed against a view of what human interests truly *are*, which inevitably draws in more comprehensive views. Gray points to Hobbes as a better thinker to work through the idea of value pluralism. Since there will be no universal consensus on the meaning of liberty or equality, or any other possible liberal values, so then must we find some other basis for unity. Hobbesian "peace" comes close to being that value. Avoiding conflict, the *summum malum* of civil association, and finding some basis for consensus is paramount; further, no one master value – not even peace itself – can do this, according to Gray. Legitimacy comes not from the value of "peace" simply so much as from a virtuous arbitration and balancing between conflicting values and views of "the good" in society, which achieves peace (2000, p.25).

There is in indeed much within *Leviathan* to suggest that Hobbes thought it unlikely that political subjects would form a consensus on matters of politics from out of their commonly shared views informed by philosophical, religious, or otherwise complex social attachments. Hobbes instead thought of these as sources of division, which Rawls captures in spirit where notes that "no other foothold" would be possible than simple self-interest for political agreement. Indeed, Hobbes broadens this view to a radical pessimism over the ability of political subjects to find some form of common political ground springing from such "comprehensive" matters, which would dampen the hope for a society united by a conception of justice in terms of anything like "public reason" or shared political values.

For Hobbes, unchecked private judgment on public matters tends only toward the destruction of the commonwealth. The "poison" of seditious doctrines, "where of every man is a judge of good and evil actions," even as rooted in matters of private *conscience*, is an ever present cause of concern for the political association (XXIX.6-7). "Pretense to (divine) inspiration," the influence of "popular individuals," and the excessive greatness of "corporations" (by which Hobbes means cities or regions or otherwise groups in civil society), are likewise all notable sources of division that can lead to a disastrous political situation (XXIX.8-21). Throughout *Leviathan* these are presented as political quandaries; they do not provide the hope of any sort of social union or political consensus, which is an impression Hobbes deepens further with some fairly deep pessimistic statements about the grounds for such consensus.

The very medium of "speech" itself, for one, and thus public judgment and "reason" become problematic for Hobbes. Unchecked, speech is the driver of endless pluralism and ultimately division. Human beings are capable of deciding what is "expedient for the common benefit," and they can *signify* this to others. This does not tend to unify persons, however. For Hobbes, people are in competition for honor and dignity thereby, and also, having the use of reason, people have the ability to find faults with the administration of their "common business" and may come to think they are "wiser" or "better able" to govern than others (XVII.7-9). Thus, while other animals can make known their desires through voice, only humans have "that art" of representing to others "that which is good in the likeness of evil and evil in the likeness of good […] discontenting men and troubling their peace at their pleasure" (XVII.10). [66]

Rawls does not fail to notice these elements, but they do not figure directly in his analysis. He tells us that Hobbes is very aware "…that he lives in an age in which people appeal to many different kinds of interests," which are religious, political, and may also, more negatively, be "interests that [Hobbes] thinks are based in the end on pride or vainglory and love of dominion" (2007, p.48). For

---

[66] Rawls notes Hobbes's study of the "classics" as being important for *Leviathan*'s method. On this point it is interesting to consider not only the history of the English Civil War, but what Hobbes may have taken from Thucydides. Hobbes gave a particularly celebrated translation of the "*History of the Peloponnesian War*"; it is interesting to consider this well-known passage regarding the consequences of the revolution in Cosyria: "Words had to change their ordinary meaning and to take that which was now given them. Reckless audacity came to be considered the courage of a loyal ally; prudent hesitation, specious cowardice; moderation was held to be a cloak for unmanliness; ability to see all sides of a question inaptness to act on any. Frantic violence, became the attribute of manliness; cautious plotting, a justifiable means of self-defence. The advocate of extreme measures was always trustworthy; his opponent a man to be suspected" (3.82.4-5). Such passages are particularly illuminating in light of Hobbes's arguments about the nature of political speech and judgement.

Rawls these do not animate the state of nature. (That is they are not needed to fuel a basic problem of mutual assurance or stability, as fear of other's unrestrained intentions of suffices for this.) They do however go a long way to explaining Hobbes's approach to "political" egoism that Rawls outlines, and that he draws toward with his an emphasis on Hobbes's language as implying a practical need for "universal" justifications. As we can see, for Hobbes the trouble is not that people are unmotivated by social or more complex interests, but that these can come to have a very problematic character.

Rawls in fact recognizes this at work in Hobbes's thought, and notes it in his lectures. Rawls sees Hobbes envisioning society as radically and irreconcilably pluralist on the subject of "the good" in general, and he echoes Hobbes from chapter fifteen of *Leviathan* here noting that even "the same person at different times will say that different things are good" just as, of course, "different people at the same time also say different things are good" (2007, p.84). Rawls also mentions that it is also not the case for Hobbes that "all in pursuing their private interests also realize the common good" (ibid). The solution being, as Rawls observes, that what "we require is some agency, some impartial arbitrator or impartial judge, to decide what is in the common good" (ibid).

Hobbes's experience of the English Civil War is raised by Rawls in reference to this, but only in his lectures on Joseph Butler. Perhaps hewing a little closer to Butler's view than his own, Rawls there describes Hobbesian subjects as nearly "unfit" for society, driven by pride and vainglory to the point that even their own reason leads them into danger, leading persons to "imagine [they] can

understand more things and run society better than anyone else"; indeed, our own use of "reason" threatens to make us into "fanatics" in this case and to "render society ungovernable, unless we recognize the situation and coolly calculate on the basis of our fundamental interest in our self-preservation" (2007, p.418).[67] If Rawls raises these points however, he does not work them into his general account of Hobbes's "political" approach.

In my next section, I will address this dynamic in terms of "stability" specifically as a political quandary. Rawls and Hobbes both see an expanded problem of stability (in terms of compliance with terms of cooperation) rooted in societal pluralism. This dynamic forms a crucial bridge (and point of difference) between Rawls and Hobbes, but Rawls does not find occasion to draw this out. It is required, however, to explain both (a) Hobbes's limited range in terms of his "free standing" core political doctrine, as Rawls describes it, and also (b) the doubts that Hobbes's thought itself raises over any suggestion that Hobbes's doctrine could serve as a point for political consensus, which will be touched upon later.

III

A problem for stability

In his lectures, Rawls says that there is no need to rely on "dramatic elements" aspects of Hobbes's theory such as *pride* or *vainglory*, as these do not

---

[67] Rawls notes that he thinks Hobbes is drawing on his experience of the English Civil War specifically with these observations (ibid).

cut towards the basic – I have also referred to it as "first" – problem of stability

that the sovereign solves (2007, p.73). Therein, political subjects need more than

just to be shown a view of what is to their mutual "interests"; they also need some

assurance that following the laws of nature will not be to their disadvantage

through exposing themselves to exploitation, which the sovereign provides by

punishing non-compliance (by making the laws of nature positive and legal laws,

and introducing coercive mechanisms). Thus, on this view, all citizens have some

reason to adhere to the "social state" as characterized by the laws of nature, as

they can be assured other citizens will not take advantage of them if they comply

with reasonable (in content) principles of social organization (2007).

This kind of basic assurance problem is also important to Rawls, as was

noted in chapter one. Rawls in fact memorably draws on Hobbes in *PL*

specifically on just this point. If we cannot rely on others to act in accordance

with reasonable principles, says Rawls, then it may become "self-sacrificial" or

"irrational" for us to act from those principles (2005, p.54). That is, "without an

established public world the reasonable may be suspended and we may be left

largely with the rational, although the reasonable always binds in *foro interno*, to

use Hobbes's phrase" (2005, p.54). [68] As Weithman (2011) notes, the basic

problem of whether we can be assured that others will 'do their part' also

animates part of *TJ*. But there it turned on whether political subjects could have

faith in other subjects as having a moral allegiance to terms of justice taking

---

[68] Reasonable terms are the domain of the "public" or shared good. Rational principles are private,
for Rawls (2005).

priority over any temptation to defect from cooperation out of self-interest, or whether the "reasonable" could be decisive over the "rational" simply.

Rawls's later political liberalism though also adds a focus on what I have called the "second" problem of stability, which has two dimensions that are both articulable through Rawls's paper "The Idea of an Overlapping Consensus" (1987). The first would be a modification of the basic assurance problem; that is, people must have an "evident intention to do their part in just or fair arrangements, [so that other people] tend to develop trust or confidence in them" (p.445). Weithman (as was noted) compellingly frames this problem in terms of pluralism. Citizens might come to suspect that the "comprehensive" groundings other people in society hold fail to provide them with the sort of reasons they require to be depended upon to "do their part." Citizens thus engage in public reason revolving around a kind of "free-standing" political conception in part to give assurance that they endorse the public conception as a method for addressing public questions of justice (2011, pp.327-35). They are thus able to endorse the legitimacy of that basic approach, even if they do not have to agree with all the decisions that are made thereby.

The other aspect of this problem is a related but more direct question. For Rawls, a political consensus is more than a *modus vivendi* when it can be expected that "shifts in the distribution of political power" would not directly threaten that conception (1987, p.433). An overlapping consensus is not a mere balancing of forces and interests; rather, it means that citizens can feel confident that no one group in civil society would use political power to enforce its "comprehensive"

conception if it got that chance. Rival allegiances to certain values must not suggest that outright defection from social cooperation is a distinct political possibility whenever the political tide shifts and the balance of power is readjusted.

In his lectures, Rawls focusses on the "basic" stability problem of assurance as the core "social" problem addressed in *Leviathan*. That is, Hobbes is said to respond to the problem of advantage seeking defection from 'cooperation' by using the sovereign to stabilize terms of political association (and thus "imposing" justice on our natures). This would not entirely address Hobbes's "political" approach however, as this work has begun to suggest. If we flesh out Hobbes's attempt to find "agent neutral" or universal reasons to justify his project, we get a little closer to the second problem of stability at work in Hobbes's thought. Hobbesian subjects have the capability to learn from "social institutions" and to have moral, or otherwise complex "social" interests, just as Rawls notes; but these do not imply any sort of possibility for such subjects to form common political ground or to be able to rely on each other not to defect from cooperation.

Indeed, if anything these sort of interests present a compelling political *problem* for Hobbes by adding an additional dimension of political instability, as we have seen. The sovereign's authority might overcome the fact that citizens do not know if one another's moral, or otherwise grounded, views give them reason to endorse the political association by replacing "moral assurance" with fear of punishments. An open question however remains over the deeper problem of outright political defection. Could "fear" of the sovereign's punishments, or

justifications of sovereign legitimacy rooted in arguments of "self-interest,"

prevent defections out of rival "comprehensive" allegiances? [69] The answer

would seem to be no, and to see why it is necessary to look into what Hobbes's

solution to the problem of political "unity" actually does. The second quandary of

stability problem calls for a much expanded view of the sovereign, as Rawls has

noted, and this chapter's subsequent subject covers.

IV

Hobbes's answer to the second problem of stability

Rawls's "moral" critique of the terms of Hobbes's agreement, as well as

the view of stability Rawls takes from Hobbes's state of nature, misses the deeper

political problem of defection rooted in "social" concerns in *Leviathan*. That is, if

political subjects were predominantly motivated by their fundamental interests,

they would simply require a sovereign to provide them with the assurance they

require that others would do their part where necessary, so that it would be

individually rational for them to do what is "mutually" rational for everyone.

---

[69] This argument is taken up convincingly by S.A. Lloyd in her *Ideals as Interests in Hobbes's Leviathan* (1992). There Lloyd argues, contra Strauss, against a political world defined by psychological egoism, where only fear of a violent death checks man's propensity to *vanity*. (That is, it Lloyd speaks contra to the classic argument that Hobbes balances one major human motivation "glory" against another "self-preservation.") Strauss had neglected *religious* pride in particular, and what Lloyd calls "transcendental interests" as highlighting the "empirically false" assertion that such interests could be checked by fear of death alone (1992, p.230-1). While Lloyd's focus on religious pride is fairly apropos, it is worth noting that Hobbes adds a number of articulations also over the trouble caused by the philosophical learning of the Greeks as well, nothing being "so dearly bought as these western parts have bought the learning of the Greek and Latin tongues" (XXI.9). Such 'learning' is something Hobbes identified as one of the causes of the English Civil War (learning in the Universities of Aristotle, Cicero, Seneca and the like [*Behemoth*, part I]).

Once this dilemma was solved the core political problem for Hobbes would also be overcome.

If this were true, Rawls might respond from one of two perspectives he raises at times. From a purely "moral" philosophical perspective, Rawls might argue that this would be undesirable as it undermines a "reasonable faith" in an authentically just society, a concern occasionally evidenced in Rawls's work.[70] Alternatively, Rawls could argue – with Gauthier – that no system of political reciprocity can be *stable* on the grounds of rational advantage alone, ignoring "moral obligation." Indeed, Rawls very arguably takes this exact line where he compares Hobbes's thought to a *modus vivendi*.[71] But either of these tracks of argument would miss addressing the core challenge of pluralism in Hobbes's *Leviathan*, for which Rawls uses language drawing towards notions of universality or commonly held interests. This is a dynamic Rawls also makes directly explicit in the passage quoted at the outset from "the Idea of an Overlapping Consensus."

---

[70] In *PL'*s section on an overlapping consensus this is retained as the Kantian idea of the "defense of a reasonable faith" (2005, p.172). There it is tied more closely into the possibility for political liberalism, but Rawls often gives this a more direct treatment in terms of whether justice "suits" our nature. It is given a stark (and again Kantian) presentation in *PL* with this thought: "if a reasonably just society that subordinates power to its aims is not possible and people are largely amoral, if not incurably cynical and self-centred, one might ask with Kant whether it is worthwhile for human beings to live on earth?" (2005, lx). This articulation appears in this exact form also in *LP* (1999, p.128). Weithman terms Rawls's project an excise in "naturalistic theodicy" in this respect (2011, p.8).

[71] In a *modus vivendi,* again, citizens appear ready to defect from social cooperation as soon as the balance of power shifts. A *modus vivendi* is not a stable social arrangement, and Rawls seems to indicate Hobbes's theory is like this where he describes *Leviathan'*s idea of the social compact as being like "to each according to their (rational) threat advantage" (2007, p.87). (This language, "threat advantage," is echoed again in terms of an undesirable view of justice in *JF* [2003, p.16]).

The more "dramatic" elements of Hobbes's work might perhaps be minimized for undergraduate lectures such as Rawls is delivering. But in terms of drawing comparisons to Hobbes and Rawls's later evolving doctrine these need to be addressed. Hobbes's 'historical' position and Rawls's own diverge significantly over the consequences of pluralism. Rawls's 'reasonable' faith, for one, requires that pluralistic subjects might find a kind of unity over terms of association that they mutually affirm. It is true that even in *TJ* Rawls does not imagine that such affirmation runs deep enough that some coercive assurances are not required to ensure, again, that political subjects will not do things like to cheat on their taxes (altruism); but later Rawls does seem to think that in matters of voting, political speech, and other public actions bearing on the public conception of justice, subjects could be relied upon to do their part.

Like Hobbes the basic gesture for later Rawls is in legitimating political power, which ultimately must enforce decisions. [72] Rawlsian subjects have a wide array of political actions they can take and thus more occasions to 'defect' from their endorsement of the shared political conception. They lend assurance in these instances by showing that they intend to "do their part" by acting in ways that demonstrate their recognition of the basic framework of public justice as legitimate. Hobbes, on the other hand, did not indicate that he thought that public

---

[72] The "principle" is summed up by Rawls in the following way: "our exercise of political power is proper only when we sincerely believe that the reasons we offer for our political action may be reasonably be accepted by other citizens as a justification for those actions" (2005, xliv). Weithman thinks *legitimacy* is crucial to *PL*'s notion of stability. An overlapping consensus on matters of justice does not entail that every political decision bearing on justice need be endorsed by all, only that the basic approach to justification is endorsed. Although this might lead to striking moments of loss for, as an example, a devout religious believer, the overall good realized by the larger political conception takes precedence.

adherence to his terms of association could be the basis for this sort of assurance; thus, Hobbes's laws of nature could not solve the deeper problem of political stability themselves by generating their own supports. Instead, Hobbes's "scientific" knowledge of the universal reasons the state serves are mobilized to endorse the legitimacy of a single *act* – the authorization of an absolute sovereign, which is what makes pluralistic subjects into a unified body.

The Hobbesian model of consensus has at its heart this act of *authorization*, which is how 'social unity' or consensus can truly be described. For Hobbes, political subjects cannot represent themselves collectively. Instead, a multitude of men are "made *one*" only when they are represented by a single "person" (XVI.13). [73] It is the "unity of the representer" and "not the unity of the represented" that makes the 'collective' person into *one*. Otherwise, unity cannot be understood in the multitude (ibid). This representation is important for Hobbes, as persons are ultimately the 'authors' of the will and actions of the sovereign collectively.

Hobbes states throughout that the sovereign is indeed to be "absolute representative of all subjects," and we have in fact all covenanted with each other (or with the representative in the case of conquest) to recognize this fact and to respect the civil law (XXII.5, XXVI.8). Again, it is this Sovereign that is the ultimate judge of doctrines and the arbitrator of all controversies (XX.3). Such authorization is arguably the single most important political act for Hobbes, and

---

[73] That is a man, woman, assembly of all, or assembly of the few.

the most crucial act of our collective 'artifice' in constructing the commonwealth. The unity it produces comes specifically through authorizing the sovereign to represent all the people. This means ultimately allowing the sovereign to serve as the judge and 'decider' between all conflicting views in society.

For Hobbes, again, man's political "being" does not naturally tend towards "social unity" in the classic sense of man as Aristotle's *Zoon Politikon* (a political animal, or natural dweller in the city or social settlements). The crucial step of artifice, the creation of the state via the recognition of sovereign authority, is a necessary first step. Diversity of opinions and judgment over how best to discharge affairs inevitably leads to dissension and conflict, as we have seen, and it is the institution of the sovereign that solves these kinds of collective action problems (XVII.4-5). Indeed, Hobbes goes so far as to flirt with a particularly dystopian solution in this respect where he suggests that the sovereign should teach "correct" political principles to the common people, whose minds are like "clean paper" – fit to be written upon if no one else has got there first (XXX.6).

This sort of an absolute sovereign cannot be explained except with reference to the problem of stability characterized by pluralism. The sovereign can solve the basic problem of stability by introducing punishments to make sure we "keep our covenants" and the like, but Hobbes felt the general justification of the sovereign as protecting our fundamental (and other) interests should also lead us to authorize a sovereign that could solve the sorts of problems his "freestanding" view might not (on its own). That is, if Hobbes had an inkling that his basic program of the laws of nature laid out the basis for a legitimate regime, he pulls

back from this conclusion over the deeper problem of political stability presented by pluralism. Hobbes's "political" argument is shaped by this reality, but it only goes so far as to endorse the idea of the 'effective' sovereign as authorized and thereby legitimate – an absolute standard of public judgment and representative of all citizens.

<div align="center">V</div>

Authorization versus an "overlapping consensus"

Rawls elsewhere tells us that we ought to consider the historical context of Hobbes's political problem. [74] Hobbes must establish that everyone has "sufficient rational grounds, rooted in their most basic interests" for creating an "effective" sovereign with "absolute powers"; and one must also consider the "civil strife" and "deep religious and political divisions" characterizing Hobbes's times, which in turn inform Hobbes's approach (2007, p.105). Rawls though never works out the contribution of a political problem of stability as informing Hobbes's 'political' doctrine and its egoism ("true enough" for Hobbes's purposes). But there is no doubt that Hobbes has read into his political "problem" a radically socially constructivist dimension to human nature, which has deep consequences for his general theory.

Consider the passage quoted previously. Human minds are like "paper" in that they can be easily written upon. However, once written upon – once subjects become "learned" or "potent" – Hobbes also goes on to tell us that they tend to

---

[74]This assertion comes at the beginning of Rawls's lectures on Locke (2007, p.105).

resist what will diminish their authority and ambition (XXX.6). Beyond such "self-focussed" motivations, we know also that pluralistic subjects compete for both honor and dignity, and that they may also come to sincerely believe they see faults in the administration of the common business that need to be corrected (XVII.7-9). These become a political quandary when married to the force of destructive *pride* for Hobbes, when they become an outright challenge to political authority, or a claim to mastery over others.

It is useful to briefly examine the role of *pride* in Hobbes's thought. For Hobbes *pride* means a vain supposing of one's power or abilities or station or worth, which could lead one to rash and unadvisable action (VI.39-41, XI.11-12). Hobbes directly identifies great vainglory as "pride" in one section which further means excessive opinion of one's own wisdom, divine inspiration, learning, or similar, which causes great excitement, vehemence, or even *rage* if not recognized (VIII.19). Consider also Hobbes's reference to it in his laws; one ought to suppose others are our equals by nature, and that no claim to be another's master by nature is likely to be accepted (XV.21). Either this is so, or at least we ought to *act* as though it is in order to avoid conflict. Any breach of this is what Hobbes calls *pride* (ibid).

Since the summation of Hobbes's laws of nature is "do not that to another which thou wouldest not have done to thyself," the significance of a basic recognition of equality is fairly salient (XV.39). Pride might be the cardinal anti-social sin for Hobbes then; but, while Hobbes thought he could make a justification to pluralistic subjects for political authority, he did not think that

same justification could be counted upon to restrain the destructive tendencies of

pluralistic subjects, as motivated by complex interests wedded to the force of

pride and other problematic aspects of human social behavior (the ability of

disturbers of the peace to convince masses of followers). For this Hobbes

introduces the idea of an *absolute* sovereign.

Compare this again with Rawls's idea of an 'overlapping consensus.' For

Rawls, citizens can represent themselves on matters of basic justice. They do so

through a process of public reason. Each recognizes one another as equals in this

sense, and, motivated by a sense of fairness, each proposes terms that they think

the other individuals could reasonably accept. [75] Like Hobbes, Rawls does not

think that any one particular view of the 'good' or what is truly worthwhile in

human affairs is ever going to 'triumph' (absent *coercion*). [76] But political

subjects are nevertheless able to have an allegiance to a free standing political

view, and they are united by a desire to deal with each other according to just

terms (2005).

---

[75] In reference to the "desire" to act fairly Rawls compares this to T.M. Scanlon (1982) and the notion of a desire to justify ourselves in terms others could not "reasonably reject" (2005 pp.49-50n). This slight turn on the language perhaps captures the idea better; it also interestingly cuts a little closer to Hobbes's own "negative" formation of the golden rule, but I will have to leave these points of argument here.

[76] Rawls does not believe in social union by way of what he calls *community*, which would be a fairly deep agreement not only on political life but also on all the purposes and ends of political life in a direct sense. A union on what Rawls terms "final ends" would entail a deep agreement on human interests, final value ends, and the means to achieve these. Here one might imagine the transcendental interest in salvation, which might require a certain religious approach to achieve (2005, p.37). Part of modern conditions for Rawls is that one such comprehensive view can only be imposed by force (ibid). In some respects it would seem Hobbes agrees, and proposes just such an 'authoritarian' solution. Yet a number of his articulations seem to support that society might need to accept a certain level of pluralism. (On the question of religion Hobbes seems to suggest a certain pluralism is a fact – "we are reduced to the independency of the primitive Christians to follow Paul or Cephas or Apollos, every man as he liketh best" [XLVII.20]. It is not clear whether Hobbes thinks this a happy development.)

Leviathan's view cannot be similarly "political" because Hobbes does not have any faith that the deeper problem of political stability can be solved except by the authorization of an absolute sovereign. Hobbes's general doctrine though *can* be politically justified in terms of fundamental human interests, which is an approach that has a 'political' shape because it responds to a particular problem of social pluralism; this is as opposed to its responding to a view of human subjects as simply "predominantly egoistic."

Whatever else we hold to, Hobbes thinks the 'effective' sovereignty can be legitimated by solving the general problem of stability the state of nature highlights particularly.[77] The sovereign thereby secures our basic interests, and, it also secures our more complex interests, too, as these must depend on our safety and means to commodious living to flourish.[78] Hence Hobbes introduces reasons that all citizens might accept, in justifying his vision of political order. There is also an assumption that such citizens will be pluralistic and divided otherwise – and their ultimate political unity can only to be finally guaranteed by authorizing an absolute sovereign to solve the problem of stability brought on by pluralism.

There can be little doubt that Hobbes regarded the sort of "reasonable" principles described by his laws of nature as "true."[79] However, Hobbes held out

---

[77] That is, again, political subjects need to be assured that they can rely on other citizens to keep their agreements, and that the law assures that by following the socially constructive laws of nature they will not be preyed upon by others.

[78] As Gabriella Slomp (2000) sharply observes, one crucial feature the state of nature has is a lack of "identity." In a state of nature a person's identity is endangered in two ways. The first is "in a crude and drastic sense: physical life is threatened"; but the second, attaching to questions of identity, comes "in a more sophisticated sense" in terms of "the distinctive ability of the mind to detach itself from the present and plan for the future," which is "wasted" in a state of nature (p.20).

[79] This goes as well for Hobbes's more general science of the state, for which Hobbes no doubt thought he had provided society a definitive knowledge. On this note, Hobbes is even given to lament whether his thoughts would nevertheless end up being as "useless" as the "commonwealth

little "political faith" in the power of his principles to win the allegiance of people generally, without some additional coercive element to enforce them (or at least teach them generally). This even though the core program of Hobbes's laws of nature – stressing equality, peace, equity, and justice – seem to be rather reasonable terms of reciprocity. Hobbes saw pluralism as radical enough that, in and of themselves, these could not be entirely politically decisive.

<div align="center">VI</div>

Consequences for Rawls's thought and conclusion

It is worth considering again a passage from the end of *Leviathan*'s chapter fifteen. Hobbes tells us that men differ not only in their "judgement on the senses of what is pleasant and unpleasant" but also significantly "what is conformable or disagreeable to reason in the actions of common life" (XV.40). This is such that "disputes, controversies, and at last war" arise (ibid). From private appetite, which judges good and evil in a state of nature, can arise one thing all men are said to agree on: that "peace is good" as are the "means of peace," the moral virtues (ibid). But if this is a universal standard Hobbes nevertheless casts much doubt upon its standing as a source for political consensus, absent coercive enforcement. With respect to stability, these can at best justify a sovereign to introduce "public" judgment over "good" and "evil," which ameliorates all the negative political effects (sources of faction) of "private" judgement.

---

of Plato" – who, Hobbes interestingly states, was "also of opinion that it is impossible for the disorders of state and change of government by civil war, ever to be taken away until sovereigns become philosophers" (XXXI.41).

Hobbes sees the "private" world in much the same way that Rawls does. For Rawls this includes "individual" judgement, but also that of private confessions of faith and communities in civil society of all kinds (2005, p.42). In short, it is everything not bounded by the "public" conception and public values. For Rawls, this milieu may nevertheless give rise to a shared world of "public reason" and also the possibility for an overlapping consensus and "freestanding" political conception. For Hobbes, civil society is a much more irreconcilable world of private associations, factions, and bodies of non-public interest generally (XXIX.8-21). The 'public world' in this context *is,* in a sense, only the sovereign itself – representative of all and standard of public judgement.

Rawls though misses out on pointing toward this 'fork in the road' between Hobbes's work and his own. Perhaps Rawls agrees with theorists such as Kavka that Hobbes's more conservative political concerns might simply be dismissed empirically as relying on "faulty assumptions" (1986, p.4). Rawls seems to do this where he calls on the work of history as vindicating, in a sense, "reasonable" pluralism. "We are the beneficiaries," Rawls tells us, "of three centuries of democratic thought and constitutional practice; and we can presume not only some public understanding of, but also some allegiance to, democratic ideals and values as realized in existing political institutions" (1987, p.422). But there is reason to be skeptical on how far Rawls can carry this point.

It is true that Hobbes's 'authoritarian' conclusion is not likely to be particularly palatable to a modern reader. However many current thinkers (Gray for one) remain very willing to maintain the viability of a *modus vivendi* based on

Hobbesian premises. Here one might also consider so-called "Hobbesian" liberalisms; indeed, a number of persuasive liberal theories might proceed according to a kind of value thin minimalism that has some affinities with Hobbes. [80] So Rawls might dismiss that an outright Hobbesian solution to modern political problems (and stability) seems suited to this historical moment, but, continuing civil strife in the world aside, even with "Western" liberal democratic societies Rawls would be hard pressed to deny some Hobbesian features in the modern political world.

There exists ample space to believe that society looks like a *modus vivendi* much of the time, with citizens accepting it more or less out of habit, or because it serves their essential interests, or because there is an awareness of state power lurking somewhere in the background. One would be hard pressed to deny that even in many "liberal democratic" societies there might be quite a lot of citizens who view themselves as "resident aliens" (to borrow a phrase from Weithman) much of time, disagreeing strongly with their society on political matters and feeling no particular ties of civic friendship or allegiance generally to the conception of justice endorsed by the society – even if such citizens are not

---

[80] Skhlar's famous "Liberalism of Fear" (1998) comes to my mind here; it emphasizes avoiding the use of government power to inflict cruelty as central to liberal politics, and it highlights toleration as a political virtue. (Skhlar would likely not appreciate the suggestion her work is "Hobbesian," however [1998, p.6]). Jean Hampton (1989) actually gives a fascinating comparison of Rawls and Hobbes on this very point. Both thinkers line up on "peace" and "stability," but for Rawls particularly "toleration" looms large in solving the problem of pluralism. Society becomes stably pluralistic through passing from a non-moral '*modus vivendi*' to a morally affirmed version of the same thing.

obviously "defecting" (yet) in any very notable way (for reasons best explained by sociological or political factors). [81]

Part of what Rawls terms a "reasonable faith" in the possibility for a just society, is the faith not only that human beings are capable of "moral" obligation and motivations, as against self-interested ones, but also that the source of these motivations can be made into a kind of "moral sense" politically. If pluralistic citizens cannot produce a shared morally affirmed conception, then what they share may end up being a "mere" *modus vivendi.* A certain political arrangement may simply coast on its inertia and ability to inspire a kind of awe; that is, until the "mortal god" of the state should one day falter and appear to all as mere hollow idol, divested of its totemistic powers.

Rawls misses a chance to compelling juxtapose his own theory with the Hobbesian conception here, particularly as it bears on the idea of public judgement. Where there is no "overlapping consensus" on matters of basic justice, the sovereign may come to look like Hobbes's towering "Leviathan," especially as it is called upon to make public judgements. In the decisive moment it would impose rather than reflect the unity of its citizens. Rawls never works out the sort of forceful defense of "reasonable" pluralism required to allow for his more 'optimistic' vision of consensus in contrast to this idea. Does a *modus vivendi* develop into an overlapping consensus naturally, just as Rawls says it does – or does it rather just develop into a still more complex version of itself, with the

---

[81] Here I echo Weithman on the idea of a *modus vivendi*, but the "reasons" provided are my own (2011, p.321).

sovereign's powers of judgement only finally accountable on the most ultimate basis: that there must be one who judges?

To sum up then, it has here been argued that Hobbes's "political egoism," such as Rawls works it out, attached not to political subjects as "predominantly egoistic" so much as their being deeply divided otherwise in their more complex social motivations. While Hobbes's core recommendations have the shape of what Rawls terms the "political," *Leviathan* does not suggest such an approach could win the allegiance of all people simply; hence, there was no really definitive answer to the problem of political stability represented by pluralism except *coercion*. Nevertheless, the view Hobbes presents – stressing reciprocal terms of equality and fair treatment – could be called reasonable, even "moral" (depending on one's appraisal). But these do not give us the 'public world' for Hobbes. Given this contrast, the deep assumption of pluralism in Rawls's later thought, and Rawls's own striking faith in the political possibility for a consensus on principles of justice for the basic structure of society, could Rawls have picked a more appropriate thinker to compare himself to than Hobbes?

Indeed, if the ability to make moral sense out of our political arrangements – to have a "reasonable faith" in the possibility for a just society – calls on Rawls to present a "realistic" view of "Utopia," then Rawls is called also to give a forceful defense of its stability on all levels. The level of "reasonable pluralism" is an area for which Rawls has given us comparatively little argument. Rawlsian theory could do very well to use Hobbes as the foil it requires to begin to make a more complex argument about the role "moral," or at least shared, political conceptions

can play on unifying pluralistic subjects. Hobbes's concerns cannot simply be banished to the shadows of history. As history itself demonstrates, the very things that troubled Hobbes – disunity, intolerance, conflict – have a habit of continually resurfacing.

References

Fisk, M. (1975). History and reason in Rawls' moral theory. In Norman Daniels, (Ed.), Reading Rawls: Critical studies of Rawls' A Theory of Justice (pp.53-80). Oxford: Blackwell.

Freeman, S. (2007). *Rawls*. London: Routledge.

Gauthier, D. (1969). *The logic of Leviathan*. Oxford: Clarendon Press.

Gray, J. (2000). *Two faces of liberalism*. New York: The New Press.

Greenawalt, K. J. (1999). On public reason. In P. Weithman (Ed.), *Reasonable pluralism* (pp.247-267). New York: Garland Pub.

Hampton, J. (1989). Should political philosophy be done without metaphysics?. *Ethics*, 99 (4), pp. 791–814.

Hedrick, T. (2010). *Rawls and Habermas*. Stanford, Calif.: Stanford University Press.

Hobbes, T., Martinich, A. & Battiste, B. (2011). A. Martinich & B. Battiste (Eds.), *Leviathan*. Peterborough, Ont.: Broadview Press.

Hobbes, T. & Seaward, P. (2010). P. Seaward (Ed.), *Behemoth, or, The Long Parliament*. Oxford: Clarendon Press.

Johnson, P. (1987). Hobbes and the wolf-man. In P. J. Johnson & C. Walton (Eds.), *Hobbes's 'science of natural justice'* (pp.139-151). Dordrecht: Springer Netherlands.

Kavka, G. (1986). *Hobbesian moral and political theory*. Princeton, N.J.: Princeton University Press.

Kukathas, C. & Pettit, P. (1990). *Rawls*. Stanford, Calif.: Stanford University Press.

Lloyd, S.A. (1992). *Ideals as Interests in Hobbes's Leviathan*. Cambridge: Cambridge University Press.

Lukes, S. (1978). *Essays in social theory*. London: Macmillan.

MacPherson, C. B. (1965). Hobbes's bourgeois man. In K. Brown & L. Strauss (Eds.), *Hobbes studies* (pp.167-183). Oxford: Blackwell.

Martinich, A. (2005). *Thomas Hobbes*. New York: Routledge.

Miller, D. (1979). *Social justice*. Oxford: Clarendon Press.

Mouffe, C. (2005). *The return of the political*. London: Verso.

Nozick, R. (1974). *Anarchy, state, and utopia*. New York: Basic Books.

Pettit, P. (1974). A theory of justice?. *Theory and Decision*, 4 (3-4), pp. 311-324.

Pogge, T. (1987). International relations as a modus vivendi. *American Society of International Law*, 81 (8-11), pp. 426–439.

Rawls, J. (1999). Justice as fairness. In S. Freeman (Ed.), *Collected papers*. Cambridge, Mass.: Harvard university press.

Rawls, J. (1999). Justice as fairness: political not metaphysical. In S. Freeman (Ed.), *Collected papers*. Cambridge, Mass.: Harvard University Press.

Rawls, J. (1999). The idea of an overlapping consensus. In S. Freeman (Ed.), *Collected papers*. Cambridge, Mass.: Harvard University Press.

Rawls, J. (1999). *The law of peoples*. Cambridge, Mass.: Harvard University Press.

Rawls, J. (2003). E. Kelly (Ed.) *Justice as fairness*. Cambridge, Mass.: Harvard University Press.

Rawls, J. (2005). *A theory of justice*. Cambridge, Mass.: Belknap Press.

Rawls, J. (2005). *Political liberalism*. New York: Columbia University Press.

Sandel, M. (1984). *Liberalism and its critics*. New York: New York University Press.

Skhlar, S. (1998). The liberalism of fear. In S. Hoffman (Ed), *Political thought and political thinkers* (pp.3-20). Chicago: University of Chicago Press.

Slomp, G. (2000). *Thomas Hobbes and the political philosophy of glory*. New York: St. Martin's Press.

Springborg, P. (1995). Hobbes's Biblical Beasts: Leviathan and Behemoth. *Political theory*, 23 (2), pp. 353–375.

Strauss, L. (1965). On the spirit of Hobbes's political philosophy. In K. Brown & L. Strauss (Eds.), *Hobbes studies* (pp.1-34). Oxford: Blackwell.

Taylor, A.E. (1965). The ethical doctrine of Hobbes. In K. Brown & L. Strauss (Eds.), *Hobbes studies* (pp.35-55). Oxford: Blackwell.

Thomas, K. (1965). The social origins of Hobbes's thought. In K. Brown & L. Strauss (Eds.), *Hobbes studies* (pp.185-236). Oxford: Blackwell.

Thucydides. and Crawley, R. (2004). *History of the Peloponnesian War*. Mineola, N.Y.: Dover Publications.

Tuck, R. (1988). Optics and skeptics. In E.Leites (Ed.), *Conscience and casuistry in early modern Europe* (pp.235-263). Oxford: Oxford University Press.

Waldron, J. (1999). Disagreements about justice. In P. Weithman (Ed.), *Reasonable pluralism* (pp.78-91). New York: Garland Pub.

Warrender, H. (1957). *The political philosophy of Hobbes*. Oxford: Clarendon Press.

Watkins, J.W.N. (1965). Philosophy and politics in Hobbes. In K. Brown & L. Strauss (Eds.), *Hobbes studies* (pp.238-262). Oxford: Blackwell.

Weithman, P. (2010). *Why political liberalism?*. Oxford: Oxford University Press.