A SELF-SERVING JUSTICE

A Critique of Rational Self-Interest at the Foundations of Social Contracts

A thesis submitted by

Simone Thibodeau

in fulfillment of the requirements for honors

in

Philosophy

Wellesley College

May 2014

Advisor:

Helena de Bres

Assistant Professor of Philosophy

Wellesley College
# Table of Contents

INTRODUCTION................................................................................................................................... 3

PART I......................................................................................................................................................... 8

THE CONTRACTARIAN PROJECT ........................................................................................................... 8

ESSENTIAL COMMONALITIES.................................................................................................................. 9

HOBBESIAN & KANTIAN POLITICAL CONTRACTARIANISM................................................................. 12

PART II .................................................................................................................................................... 24

THE NORMATIVITY OBJECTION .............................................................................................................. 24

PART III................................................................................................................................................... 37

POWER, INABILITY, & PROXIMITY............................................................................................................ 37

CHAPTER I: POWER AND ADVANTAGE.................................................................................................. 42

CHAPTER II: INABILITY ............................................................................................................................ 54

CHAPTER III: OUTSIDE THE CONTRACT................................................................................................. 64

CONCLUSION .......................................................................................................................................... 72

BIBLIOGRAPHY ....................................................................................................................................... 76
Introduction

Modern Western Philosophy has claimed to establish universal concepts of social and political justice that transcend societies and individuals. At the same time, it has tended to base its view on the self-interest and reasoning of the individual alone. Many of the best philosophical minds have sought to show that these two assertions are compatible by showing that social justice can be established on the basis of individuals’ mutual self-interest. In other words, the good of the whole can be deduced from the interests of the individual.

However, in common understanding, even within the political sphere, justice and self-interestedness seem little connected, and often at odds. Selflessness and altruism are not only commonly featured virtues in moral theory, but are common cultural, religious, and political values. While most modern western political theory presumes there to be some separation between comprehensive moral understandings of the world and the policy decisions we must all mutually abide by, a perceived threat to political (and not simply moral) justice is recognized in rational self-interest. Political, philosophical, psychological and economic theories struggle with the collective action problems societies will face should they have to negotiate policy between rationally self-interested individuals, who abide only by their own perceived advantage rather than by some other moral code. The tragedy of the commons, the free rider problem, and other recent unsolved problems facing both academic and practical realms arise from models of governance based on the presumption of self-interest as a foundational principle. It is little wonder that in common political discourse - whether inspired by liberal or
conservative viewpoints - politicians, political parties and even politics itself are readily denounced and accused of allowing self-interested ambition to supplant justice.

The virtues are lost in self-interest as rivers are lost in the sea.

-- Franklin D. Roosevelt, 1933

However [political parties] may now and then answer popular ends, they are likely in the course of time and things, to become potent engines, by which cunning, ambitious, and unprincipled men will be enabled to subvert the power of the people and to usurp for themselves the reins of government, destroying afterwards the very engines which have lifted them to unjust dominion.

-- George Washington, Farewell Address, 1796

However, when examining the dominant political theory of contemporary western thought, that of social contract theory, we see that, rather than playing a villainous role, self-interest is supposed to be at the foundation of justice. For some philosophers, justice is merely a by-product of individuals’ natural right to pursue their own self-interest, while for others self-interest plays a more subdued, but nonetheless fundamental role in the establishment of a just social order.
As this paper will seek to show, despite these theories’ best efforts to establish a universal basis for social and political justice on the interests of self-serving individuals, problems remain and each conception of social contract theory falls short of establishing the theoretical political justice they seek. Different versions of social contract theory each offer a solution to the tension between the interests of the individual and the interests of some or all other individuals. Many attempts to bridge these two perspectives have been made. Some are more convincing than others, but all leave unavoidable gaps in their account of justice, either due to their inability to reconcile the individuals’ self-interest with the laws those individuals agree to follow, or because of the consequences of discrimination, exclusion, or differential treatment that a political contract between self-interested agents would permit.

In this thesis, I examine the role that self-interest is supposed to play in the creation of political contractarian justice\(^1\), and flesh out the consequences of its involvement. I explore the degree to which self-interest serves as the foundation and justification for political justice in these theories, and whether or not that role proves to be a weak point. Given the great philosophical variations within political contractarian theories, it would be misleading to have a single method of critique. At the same time, surveying every social contract theory ever devised might obscure the emphasis on objections to the common structure. Therefore, I will narrow my focus to the Hobbesian

---
\(^1\) The terms “contractualism” and “contractarianism” are sometimes used to refer to different camps within social contract theory. Contractarianism often refers to Hobbesian social contract theories, while contractualism refers to Kantian versions (Cudd 1). However, to have a term that encompasses both camps, I will use the word “contractarianism” broadly, referring to all contract based political theories. I will indicate when I intend to limit my reference to one camp or the other by designating a contractarianism “Hobbesian” or “Kantian,” respectively.
and Kantian camps of Political contractarianism, as these two are often perceived as the principal dichotomy in social contract theory (Southwood 2008 184).

Within these traditions, social contract theorists differ according to which part of political justice they are concerned with. Some use the contract device to answer the question of political legitimacy ("what, if anything, can legitimate the rule of some over others?"). Others use the contract device to answer the question of political obligation ("under what conditions are we obliged to obey?"). Still others use the device to answer the question of justice ("which principles would be appropriate for fairly regulating the basic structure of society?"). However, in each case, the answer involves the idea of consent on the part of a set of "parties" to a (actual or hypothetical) contract. This consent is either entirely or partially based on personal interest, rather than (purely) an interest in preserving justice. While the projects of different contractarian political theories may vary, their methodology and reliance on self-interest is shared. Therefore, the following question can be meaningfully addressed to all of them: is the social contract at the foundation of political morality fundamentally grounded in self-interest? If so, is such a contract capable of establishing a just social order?

As I wade through these differences within political contractarianism, I will compare how self-interest is incorporated in each conception’s foundations, as well as examine the different effects that self-interest’s many roles can have on resultant doctrines of political justice. In Part I, I will begin by giving a rough sketch of the essential features of political contractarianism, and then describe the important differences between Hobbesian and Kantian camps. In Part II, I will explain how using self-interest as an exclusive grounding for political justice may lead to a self-
contradicting, unenforceable legal code. As Hobbesianism relies more heavily on rational self-interest for its foundation, the issues this part raises will be mostly troubling for that side of political contractarianism. Finally, in Part III, the consequences of the particular stipulations of a justice founded on self-interest will become more evident. As I will explain, the problems of exclusion and potential discrimination necessarily ensuing from the incorporation of self-interest as a key component of justice, without some sort of independent and overriding code of justice, makes it impossible for either Hobbesian or Kantian theories to provide a basis for an inclusive, universal political justice.
Part I

The Contractarian Project
Essential Commonalities

Before delving into the variations within political contractarian theories, it is important to sketch out what the school of thought usually entails. While conceptions of contractarianism vary dramatically, there are some shared themes that distinguish it from earlier or competing theories of governance – such as divine right, theocracy, etc. Such commonalities include the requirement of some measure of acceptance, consent, or free choice of the governed, the need for hypothetical, rather than actual historical instances, of contracting, and a role for self-interest. Ideally, social contract theory finds a basis for political legitimacy in the agreement of free, rational and equal persons (Lloyd and Sreedhar 2014).

For our purposes, as we are exploring a political contract, not a moral one, we might want to think of the project as balancing the need for control with the need for certain freedoms, all with individual self interest in mind. As all of these things conflict, and cannot be protected entirely, comprehensively, easily or perfectly, the need for agreement arises – a need which motivates the creation of a social contract. But on what basis is such an agreement legitimate? Typically in contractarian political theories, agreement or free consent is the basis for legitimacy, which means that the motivations of persons will determine the existence and nature of a social contract. What drives this consent is debatable, but in all social contract theories it is, at least in part, driven by self-interest.

Having an agreement about what societal code one can be required to live by respects freedom, by allowing those subject to the code to decide the conditions to which
they are subject, while also allowing for constraints by establishing a rule of law. As Nagel describes it, the goal for social contract theory is to create a contract of which it can be said that:

No one will have grounds for moral complaint about the way it takes into account and weighs his interests and point of view. Even though be may be able to think of alternative arrangements more advantageous to him, still on balance, taking everyone’s point of view into account together with his own, no one living under such a system will have grounds for objection to the way it treats him.


This requirement of acceptance by the governed is expressed differently in other social contract theories. However, a provision of acceptance of some kind is universal to contractarian theories.

However, as there is no practical way to gain the continuous, expressed consent of all people in each generation and still to maintain that rule of law, this agreement cannot be a literal one, but has to be hypothetical. Thereby, we have to make an assumption on what people’s motivations in the contractarian position are. Almost universally in social contract theory, one of these assumed motives determining consent and thereby greatly affecting the social contract is self-interest. Using self-interest as a foundational element to a social contract initially appears to imply that:
1) People are, as a matter of psychological fact, fundamentally self-interested

2) Self-interested motives of persons are a morally rightful source of political legitimacy,

   and/or that

3) Self-interested motives will lead to the creation of a just society, in that the society is based on the sort of contract everyone subject to it would agree to.

All of these claims are contentious, but, for the sake of focusing on political philosophy, rather than psychology or moral philosophy, I will focus on the third. I will not touch on common criticisms of the social contract by refuting psychological egoism. It is plausible that most people are not psychological egoists (i.e. exclusively motivated by self-interest). This means that the psychological assumptions built into social contract theory can conflict with the moral and reasonable sensibilities of actual persons. This is of crucial importance, because, to be justified, a social contract must meet the full publicity condition, meaning that the social contract must be acceptable to members of a well-ordered society (Cudd). This is a serious problem for some social contract theories, but it will only factor into this paper indirectly.
As my main critique of contractarianism will surround its use of rational self-interest, the Kantian-Hobbesian divide is a natural one to explore, since the theories on either side of this divide have very differing accounts of how rational self-interest is relevant to social contract theory. While I intend to focus exclusively on political contractarianism, many of the modern Hobbesian contractarians I will reference will be discussing moral contractarianism. David Gauthier, Jean Hampton, and Gregory Kavka all focus either partially or entirely on moral contractarianism. Nonetheless, some of these authors’ works clarify or elaborate on Hobbes important ways that are relevant to political social contracts. I will not critique each of their theories in full, but rather will use relevant parts of their work to defend and understand Hobbes’ contractarianism more fully.

The original sixteenth century Hobbesian social contract is now a well-known formula. I will discuss some of its tenets in depth in parts II & III. However, it will be useful to give a condensed sketch of it here. Before the existence of government, rational self-interest\(^2\) and fear reign supreme in human motivation. Therefore, persons are naturally free to pursue their own ends as they best see fit. Hobbes refers to this freedom as the Right of Nature (I.XIV.i). Additionally, all individuals are supposed to be roughly equal, in that they each have the capacity to be a lethal threat to each other, given the right opportunity (I.XIII.i). Being roughly equal, individuals will hope to be able to attain

\(^2\) Hobbes refers to self-interest as “reason.” I will avoid this terminology because of potential confusion with Rawls’ understanding of the difference between rationality and reasonability.
similar resources, over which they will conflict, instigating the war of all against all. As there is no rule of law to assure individuals that agreements will be honored, there can be no cooperation between them. They will seek to destroy each other, because they fear each other’s roughly equal power to harm (I.XIII.iii). No justice, which Hobbes defines as obedience to law, can precede a coercive state (I.XIII.vi). Persons must transfer their rights to a sovereign before their freedom can be limited.

Unfortunately, this combination of individual self-interest, freedom, and rough equality without some sort of coercive state to define the limits of what human beings may do in the pursuit of their own self-interest creates problems of collective action and general mistrust that lead to the disastrous state of nature, which Hobbes describes in the most famous and often, quoted passage of *Leviathan*:

> In such condition, there is no place for Industry; because the fruit thereof is uncertain; and consequently no Culture of the Earth; no Navigation, nor use of the commodities that may be imported by Sea; no commodious Building; no instruments of moving, and removing such things as require much force; no Knowledge of the face of the Earth; no account of Time; no Arts; no Letters; no Society; and which is worst of all, continuall feare, and danger of violent death; And the life of man, solitary, poore, nasty, brutish, and short.

-- Thomas Hobbes, *Leviathan*, Part I, Chapter XIII, Section v
To escape this vicious trap, individuals must transfer the natural rights and abilities that define them within the state of nature - unmitigated pursuit of self-interest, and rough equality of lethality - to a sovereign, who will establish the rule of law (I.XIII.ii). Their rationally self-interested nature will compel them to give up these freedoms to a sufficiently powerful sovereign so as to escape the horrible state of nature in favor of the preferable state of peace (I.XIII.vi, I.XIV.iii-v).

Concerning Kantian contractarianism, I will largely focus on the theory put forth by John Rawls in *A Theory of Justice*. There are two main structural differences between Hobbes’ and Rawls’ account, one pertaining to the nature of the contracting situation, the other concerning the motivations of those contracting. Rather than contracting from a state of nature in which persons are fully knowledgeable of their situation, Rawls’ “original position” limits people’s awareness when contracting, so they cannot attempt to create a contract that unfairly caters to them in their social position (a social position which, itself not created by a fair social contract, cannot be presumed to be just). I will discuss the reasons for a veil of ignorance as well as Rawls’ understanding of what societal facts are morally arbitrary in depth in Part III. For now, suffice it to say that by limiting knowledge of one’s position within society, the veil of ignorance is supposed to provide a contracting position in which persons are truly equal.

Rawls readily admits that the original position is a hypothetical situation that cannot be expected to ever exist or to even be possible: “This original position is not, of course, thought of as an actual historical state of affairs [or even a possible one] . . . [but] our social situation is just if it is such that by this sequence of hypothetical agreements we would have contracted into the general system of rules which defines it” (Rawls, 540).
The validity of referring to the original position and the veil of ignorance, then, is that it is the closest we can come to modeling an original fair state of affairs, which is the only basis from which we can form a social contract that could be considered fair and just. As we can clearly conceive of the original position, we can thereby discover what we would consider just if we were not biased by the injustices of our existing basic structure of society. Since the position of the original contractor is free of bias and all morally irrelevant considerations, this perspective is preferable from the standpoint of justice to any other, or so the argument goes.

But first, why should we care about the opinion of those in the original position if it is not based on the pure maximization of our self-interest, given our positions in society, as is Hobbesian contractarianism? Rawls is ready with a straightforward and largely convincing answer: we already do. The opinions of those in the original position are necessarily valuable to us:

[T]he conditions embodied in the original position are ones that we do in fact accept. Or if we do not, then perhaps we can be persuaded by philosophical reflection. Each aspect of the contractual situation can be given supporting grounds. Thus what we shall do is collect together into one conception a number of conditions on principles that we are ready upon due consideration to recognize as reasonable.

While contractors behind the veil of ignorance are limited in their knowledge of their particular interests, there are many individual interests they will seek to secure and promote in a social contract. They will know that they wish to accomplish something in life, through a career, a hobby, advocacy or some sort of other constructive pursuit (although they won’t know which particular version(s) of such projects they will favor). Also, they will be aware that they, as persons in a society, are likely to value certain personal relationships, whether with friends, family, romantic partners, etc or group associations, such as religious and political affiliations. Given their awareness that they will, outside of the veil of ignorance, have a “Rational life plan” including but not limited to these things, they are likely to promote a contract that would allow them, whomever they are outside of the original position, to pursue such a plan. When choosing principles of justice, this is the sort of rational self-interest they will be employing (Freeman, Rawls 2007, 148).

While contractors are not as knowledgeable in the original position as they would be in the state of nature, they are not entirely ignorant of the world around them, and have a basic understanding of the facts of life and social interaction: “[The contractors] understand political affairs and the principles of economic theory; they know the basis of social organization and the laws of human psychology” (John Rawls, A Theory of Justice, 137).³

³ As Deborah Kearns rightly says, “I can hardly imagine four more debatable topics.” (Kearns A Theory of Justice-and Love, 37). There are a great series of objections to Rawls’ understanding of moral psychology, which do have bearing on the legitimacy of the theory. However, as I have decided to refrain from exploring social contracts from a psychological perspective, I have not included such a critique in this thesis. It may very well be that Rawls’ theory of justice necessitates a good deal of moral assumptions (which are themselves, in turn, subject to philosophical scrutiny), despite his statements to the contrary in his later works, such as Political Liberalism. While Rawls himself may not have been able to stand by such a result, making moral and psychological assumptions that value fairness does not, in itself, invalidate Kantian Contractarianism – just as assumption of ethical egoism
I have explained so far that those in the original position are not only guided by general facts about the world, but also by a certain kind of rational self-interest and a pre-existing desire to create a just social contract, one that all contractors could agree to. In addition, and importantly, Rawls’ agents behind the veil of ignorance are not only rational but also reasonable. Both rationality and reasonability impose certain requirements on the deliberation process in the initial position, but are nonetheless distinct. While rationality pursues the good for a person, which, behind the veil of ignorance largely consists of maximizing the minimum position, thereby creating fair equality of opportunity and securing liberty to pursue different conceptions of the good, reasonableness includes certain ideas about “rightness,” the requirements and duties imposed upon individuals and institutions in a just society, and a sense of justice (Freeman 150). A “reasonable” person is more inclined toward cooperation, and making concessions of his rational self-interest for the sake of meeting certain moral demands (Freeman 150).

Such reasonableness is a faculty that, according to Rawls, humans develop along with rationality:

Just as persons gradually formulate rational plans of life that answer to their deeper interests, so they come to know the derivation of moral precepts and ideals from the principles that they would accept in an initial situation of equality.

---
does not, in itself, invalidate Hobbesian contractarianism. Both theories have underlying conceptions of human moral psychology and morality itself. Whether these ideas concerning morality and psychology are themselves correct bears greatly on the correctness of these political theories. However, having moral and psychological assumptions does not, in itself, invalidate them as political theories.
Ethical norms are no longer experienced merely as constraints, but are tied together into one coherent conception.

-- Samuel Freeman, *Rawls*, 433

Because of Rawls’ use of peoples’ “sense of justice” as a justification for the original position, it is important to note that he assumes that all those in the original position, will have already developed, and will continue to be motivated by, a desire to promote this sense of justice in themselves. As he puts it, “Obviously the purpose of these conditions is to represent equality between human beings as moral persons, as creatures having a conception of the good and capable of a sense of justice” (547). While Rawls contends that his theory of justice is inextricably linked with the theory of rational choice (Rawls 547), rational choice alone is not seen to be the sole motivation of persons that needs to be considered when forming a social contract. This justice as fairness, upon which this sense of justice is based, is independently valuable and needs to be preserved in a social contracting position.

This fundamentally contradicts the claims of Thomas Hobbes and neo-Hobbesians, whose contention that justice cannot exist outside of the state of nature is based on the idea that there is no independent justice that precedes a situation in which acting according to law maximizes individuals’ self-interest. Justice only exists in a

---

4 Please note that a conception of the good is very different from a sense of justice, as explored in Part I. In this section, I will address exclusive people’s senses of justice, as conceptions of good are hidden behind the veil of ignorance and do not thereby contribute to the possible logical fallacy I will explore here.
situation where the establishment of a sense of justice is rationally maximizing for contractors, and is not an independent value, separate from rational choice:

And therefore where there is no Own, that is, no Propriety, there is no Injustice; and where there is no coercive Power erected, that is, where there is no Common-wealth, there is no Propriety; all men having Right to all things: Therefore where there is no Common-wealth, there nothing is Unjust. So that the nature of Justice, consisteth in keeping of valid Covenants: but the Validity of Covenants begins not but with the Constitution of a Civill Power, sufficient to compell men to keep them: And then it is also that Propriety begins.

-- Thomas Hobbes, *Leviathan*, Part I, Chapter XV, Section II

Moral constraints arising from what are, in the fullest sense⁵, conditions of mutual advantage . . . [and] the disposition to comply with moral constraints, without which moral relationships fall victim to the scorn of the Foole, may be rationally defended only within the scope of expected benefit.

-- David Gauthier, *Morals by Agreement*, 268

---

⁵ By "the fullest sense" Gauthier here means "exclusively"
This puts the neo-Hobbesian and Rawlsian conceptions of justice at radical odds. While Rawls intends to employ rationality, or individual self-interest, for the sake of the promotion of justice as well, Hobbes sees justice and political legitimacy as grounded in the exclusive pursuit of self-interest. For Rawls, self-interest is not used because of some value placed on “selfishness” or pure, unmitigated egoism, but for the reason of making sure all persons behind the veil of ignorance are represented equally, so their needs are fairly catered to (Freeman, Rawls 2007, 151). If people’s interests were to vary, not only including their personal interests but the interests for others that they care for, there might be an unequal representation of certain parties, whose interests happen to be in mind of multiple parties. Conversely, if a person is more isolated, and is the singular (or the one of few) people who value his interests, he might not be equally accounted for in the social contract. This phenomenon of the inequality produced by considering more than one’s own self interest within the initial position is often referred to as “double counting.”

Participants regard themselves as being as equal as all other agents, and therefore equally deserving of a good life in the societal structure to be created. Rationality for Rawls is to be conceived of as the ability to access “the most effective means to given ends” – those ends being one’s self-interest. Because of their commitment to their own equality, contractors do not think about justice in terms of what is a net good for the whole of the population. Instead, they are concerned with securing their own rights and their ability to pursue their own conception of the good (whatever that turns out to be outside of the veil of ignorance), rather than maximizing net or average welfare for the whole. For example, in the original position, contractors would not be inclined to adopt utilitarianism, as it would justify allowing great inequalities and suffering among some as
long as it produced a net “good” when compared to the prosperity in other regions of society (Rawls, A Theory of Justice: Revised Edition 1999, 13).

So, Rawls assumes that those behind the veil of ignorance cannot “rig” the social contract to their particular advantage, while disadvantaging others. This has to do with their limits of knowledge, so that, once in the contracting position, agents cannot have particular desires based on their personal advantage given their position within an existing social structure – which may itself be unjust. If they advantage some parties at the expense of others, they risk themselves being a member of the disadvantaged group upon the lifting of the veil. This allows them to be impartial, and to consider how individuals from all parts of society will be affected by the social contract, because they have no special interest in any one agent. Given they are unaware of the positions each of them might assume in the society they are positing, and that they are unattached to other agents, an individual contractor will not have any unjust bias, and will have no rational reason to attempt to make certain individuals much better off than others (Freeman 2). Thereby, they are not selfish agents bent on domination of others, in competition with others, or desirous of exclusive advantage, as we might consider Hobbesian contractors to be in the state of nature. Instead Rawlsian contractors are self-interested in the sense that they are “mutually disinterested” in each other’s interests - this to prevent double counting and make sure all contractor’s interests are accounted for equally (Rawls, A Theory of Justice: Revised Edition 1999, 12).

With these motives and this basis of knowledge set in place, Rawls finally begins establishing principles of justice based on how these hypothetical, rational, morally capable, mutually disinterested, and strategically ignorant agents would choose.
Both of Rawls’ resulting principles of justice, which constitute the basic social contract, emphasize equality. The first principle is of the highest lexical priority to Rawls, and appears uncontroversial in its claim: that a satisfactory and equal scheme of basic rights must be insured to all, and their “fair value,” or ability to be meaningfully, effectively and equally used in society, must be guaranteed to each individual by providing the means to exercise rights. Basic rights and liberties in Rawls’ view consist of political liberties such as free speech and assembly and the right to vote, freedom of religion and thought, freedom and integrity of the person, a right to hold property, and finally, rights covered by the rule of law, such as the right to a fair and speedy trial (Pogge, John Rawls: His Life and Theory of Justice 2007, 82-83). These all seem consistent with common moral beliefs outside of the veil, but Rawls believes the principle is so essential that those in the original position would unanimously choose it. The secondary rule of justice further regulates the distributions of rights, liberties, and the means and resources to exercise them. Rawls, in his latest articulation of this principle, states “Social and economic inequalities are to satisfy two conditions: first, they are to be attached to positions and offices open to all under conditions of fair equality of opportunity; and second they are to be to the greatest benefit of the least advantaged members of society.”

In what follows I will not specifically critique the content of the resultant principles of Rawlsianism, as equivalent principles are not readily evident in Hobbesianism, which assumes that any social contract with a coercive sovereign who is truly powerful enough to be coercive and to protect his subjects from an outside power, is

---

preferable to the state of nature, and is therefore rationally imperative to support (Hobbes I.XV.iii). However, there are some stipulations in Hobbesian contractarianism about government power under an established sovereign that are important to note. Most notably, justice established under a social contract cannot require that individuals act so contrary to their self-interest that they go entirely against their nature. While persons’ ability to act in their self-interest can be limited by a sovereign, a sovereign cannot require that individuals cease to be self-interested. This means that individuals cannot be required by justice to submit willingly to execution or incarceration, to testify against oneself, etc (Hobbes 2012 [1651], I.XIV). This will be examined at length in the following section.
Part II

The Normativity Objection
The validity of Hobbesian contractarianism relies on the truth of the assertion that, once a social contract is made, self-interested persons will find it rational and maximally self-serving to obey and follow the agreements they have made, without any supplementary moral principle – such as that, one ought, regardless of benefit, to abide by one’s promises (Cudd 2012, 1). To avoid an unnecessary foray into moral psychology, I will not discuss what actual persons might do within a society, and will not reference any actual violations of social contracts for individual benefit. I will merely examine whether, given Hobbes premises about the foundations of political legitimacy, an effective Hobbesian contract would even be possible.

Hobbesian hypothetical or counterfactual agreements are not morally binding if they do not follow the motive that bound them: self-interest. Hobbes does not allow for any rules, sense of justice, or codes of political morality outside of what is demanded by self-interest. Thereby, adhering to social contracts must be individually beneficial to citizens, either because of the benefits of cooperation and society, or because of the potential punishments of disobedience. Should obedience to the law not be defensible by rationality, particularly in specific instances when breaking that law would benefit the rational agent, the entire project of social contracts based on self-interests fails. As neo-Hobbesian David Gauthier puts it,

Indeed, if our defense fails, then we must conclude that a rational morality [or, for our purposes, a rational social contract] is a chimera, so that there is no rational and impartial constraint on the pursuit of individual utility.
Hobbes himself acknowledged this potential problem, in his discussion of the Foole’s objection. As Hobbes posits him, the Foole asserts that justice, which by Hobbes’ definition is simply obedience to a social contract, is sometimes contrary to reason, which is man’s method of discerning his own good. Individual good is the justification for the very social contract that persons are compelled to obey. Therefore, there is no imperative of justice to require that an individual obey the laws to which he has agreed when it is in his self-interest to break them, “and if it [disobedience] be not against Reason, it is not against Justice; or else Justice is not to be approved for good’” (Hobbes 2012 [1651], I.XV.iii).

In game theory, which is often associated with neo-Hobbesianism and is a main source of inspiration and reference for David Gauthier, Gregory Kavka, and Jean Hampton, this problem is also formulated as the free rider problem. In such a model, (often illustrated by someone taking public transit without paying for that service) goods and services are taken advantage of without paying for or supporting the infrastructure that provides these goods and services. As the free-rider formulation of this problem is limited in scope, I will not address it separately in this paper. However, it is important to note that many of the philosophers and Hobbes-inspired economists I will refer to in this paper refer to the normativity objection as the free rider problem. I object to this usage, as it excludes rational violations of the social contract that have nothing to do with economy or transfer of goods and services, but the free-rider problem can be seen as one part of the normativity objection.
However, Hobbes finds this reasoning specious and false, though he does admit the format of the question as being legitimate: “where one of the parties has performed already; or where there is a Power to make him performe; there is the question whether it be against reason, that is, against the benefit of the other to performe, or not” (ibid). Hobbes asserts that while this could be a concern, it is in fact in accordance with self-interest to follow through with one’s commitments and obey a social contract. He justifies this assertion with some speculations about the unpredictability of the future, and the uncertainty of the consequences of one’s actions.

And I say it is not against reason. For the manifestation whereof, we are to consider; First, that when a man doth a thing, which notwithstanding any thing can be foreseen, and reckoned on, tendeth to his own destruction, howsoever some accident which he could not expect, arriving may turne it to his benefit; yet such events do not make it reasonably or wisely done.

-- Thomas Hobbes, Leviathan, I.XV.iii

From this passage, we can see that Hobbes finds the breaking of covenants unilaterally hazardous, even in situations where we believe we can escape the negative consequences of such action. It also implies that Hobbes’ idea of what is correct and “wise” is not entirely determined by the ultimate consequences, because such things are unforeseen. Uncertainty is especially important when one considers the ultimate negative
consequence Hobbes envisions for the man who breaks the social contract: expulsion from society, and return to the state of nature. Any risk so potentially destructive to one’s own life is prohibited by reason (Hobbes 2012 [1651], 68).

Some neo-Hobbesians are satisfied and convinced by this argument, and have attempted to defend it from critique. Gregory S Kavka, for example, asserts that Hobbes does not need to deny that violations of contract that turn out to maximize self interest after the fact and with absolute certainty are in accordance with reason (though Hobbes does in fact do this), but only that, before all the consequences are known, such high costs of error make such a risk unreasonable.

As Kavka understands Hobbes, his position asserts a rule-egoistic system, creating a set of rules that will yield better outcomes by following than by evaluating one’s self interest on a case-by-case basis. However, Kavka notes that even Hobbes’ laws of nature have “escape clauses” which allow for certain violation of these laws in the face of the unreasonableness of certain demands. For example, Hobbes permits “defensive” violations of rules, meaning that one may break with agreements or violate other laws of nature when the alternative is being taken advantage of by others who violate the laws (Kavka 1995, 8). We can see added grounds for Kavka’s assertion in Hobbes’ claim that no one can be required to not defend his own life, even against the state and social contract that supposedly binds him. This includes imprisonment, death, testimony to one’s own guilt, etc (Hobbes 2012 [1651], I.XIV). In addition, Hobbes admits that calling these principles "laws" is misleading, for the "laws of nature" are simply conclusions drawn from natural reason (the pursuit of self-interest) rather than as actual
mandates, either inherently forceful or derived from an authority (Hobbes 2012 [1651], Hobbes I.XV.iii).\(^7\)

However, many neo-Hobbesians find this response to the Foole unsatisfying. David Gauthier and Jean Hampton, while they agree with Kavka that a principle case-by-case maximization of utility would leave Hobbes in a weak position to counter the Foole’s objection, have taken different revisionist approaches, believing that rule-egoism and the motivation of risk aversion are not sufficient to make rule-violation irrational (Kavka 1995, 13). As David Gauthier mentions, the Foole might accept that there are risks that make it usually rational to follow laws and promises (Gauthier, The Logic of Leviathan: The Moral and Political Theory of Thomas Hobbes 1969, 165). However, this does not change the fact that if there were a situation in which it were either a small enough venality, or a large enough reward, it would still be personally, individually rational to violate one’s agreements. To counter the concerns of the Foole and support Hobbes’ ideas about uncertainty and risk, some Hobbesian contractarians, including Gauthier have insisted that limiting one’s pursuit of self-interest under a social contract, even in particular situations where this will keep one from maximizing one’s own self interest, is itself rational.

Gauthier argues that the rational Hobbesian employs so-called “constrained maximization.” Genuine, inviolable obligation arises from the formation of a rationally self-interested covenant or social contract, and this forces individuals to give up their right to be purely rational egoists once within society. Any “natural right” to violate

---

\(^7\) Though, he does conclude that it is the correct term in the theological, sense, as they are commanded by god: “But yet if we consider the same Theoremes, as delivered in the word of God, that by right commandeth all things; then are they properly called Lawes.” Hobbes, Thomas. Leviathan (pp. 74-75).
collective law, when it is given up for rational reasons, is thereby irrational to reclaim, even when doing so might maximize one’s utility. This is a definitional truth considering Gauthier’s account of “obligation”: “‘A has an obligation not to do X’ = A has laid down the natural right to do X, as a condition of preservation” (Gauthier, The Logic of Leviathan: The Moral and Political Theory of Thomas Hobbes 1969, 58). This might seem similar to Kavka’s rule-egoistic system, but is actually stricter. According to rule egoism, any type of action that overall will lead to our better welfare is rational and therefore imperative to follow. In contrast, in Gauthier’s understanding, laying down our right to exclusively pursue our individual good is given up for the sake of that rational good (to form a contract, and be free of a state of nature). Therefore, not even a rule-egoist calculus of whether to obey or not obey laws would be permitted, as our right to consider such things and undergo such a self-interested calculus is now irrational to employ at all. In this way, one’s right to maximize one’s self interest is absolutely constrained by these rules he had laid down to preserve that very self-interest (Kraus 267).

Now, this may, at first, appear to be begging the question the Foole and free-rider problems intend to pose. Gauthier presumes that the renunciation or transfer of certain rights (other than the right to protect one’s immediate preservation) is theoretically possible, and thereby unconditional obligation is possible (Gauthier, The Logic of Leviathan: The Moral and Political Theory of Thomas Hobbes 1969, 58). This is a substantial psychological claim: that one can rationally give away one’s rational ability to consider what is in their rational self-interest. In obligating oneself, one rationally agrees to be, to some degree, irrational:
Hobbes supposes that all obligations are undertaken on prudential grounds—as a means to self-preservation. But this does not make obligation a prudential concept. To say that obligations are undertaken for prudential reasons is not to say that we are under obligation only so long as we can find it to our self-interest.


While this ambitious claim might, if possible, divert the critique of the normativity objection from Hobbesian contractarianism, it is not clear how such a rational renunciation of rationality is consistent with the rest of Hobbesian contractarianism. Even if it were theoretically possible to lose one’s capacities to consider one’s individual rational self-interest upon agreement to a social contract, there is no rationally self-interested reason to alienate it entirely. In truth, there is no total alienation from rational calculation within a Gauthier-Hobbesian system, as there is a similar escape clause to obedience when such obedience threatens the self-preservation of the individual. In addition, one is allowed and encouraged to reason rationally within the confines of the rules of a Hobbesian society. However, this right to reason rationally ends where those rules begin, unless those rules threaten your life. This is rather arbitrary from

---

8 This possibility is more relevant to psychological rather than political theory. I will, thereby, refrain from critiquing it in this thesis.
a Hobbesian perspective, as the reason for the erection of such rules in the first place is such a rational calculus.

Gauthier himself sees how Hobbes’ fool might not see minimax relative concessions\(^9\) as a satisfactory answer, and appeals to Kavka’s additional idea about the uncertainty of circumstances making following certain rules rational to follow even when it seems that in certain individual cases it might profit the actor to violate that rule (Gauthier, Morals by Agreement 2006, 165). However, Gauthier would assert that this would make rule following rational before \textit{and after} learning the exact consequences of an action, having committed that action oneself. This is where he and Kavka diverge.

Let us grant for a moment that the psychological feat of making oneself irrational is possible. Aside from this, even if it were possible to give up one’s rational abilities, it is not clear how the subsequent irrationality can be justified by the standard of rationality upon which Hobbesianism is based. Otherwise stated, even if we could make ourselves irrational to comply with an agreement that it is rational to make, how can it be in itself rational to do so? Gregory Kavka explains what he perceives as the absurdity of the Gauthier-Hobbesian position by posing what he calls “the toxin puzzle”:

\begin{quote}
An eccentric billionaire offers you one million dollars if, at midnight tonight, you will sincerely promise to drink at noon tomorrow a glass of toxin that would make you very sick for a
\end{quote}

\textsuperscript{9}This is the term Gauthier uses to describe the end to which rational, constrained will aim. In Gauthier’s social contract bargaining position, agents will seek to maximize their minimum standing (ensuring the best conditions for themselves even in the worst of circumstances) with some reference to what other agents are given (as they will not feel that their rational self-interest is maximally satisfied by the social contract if the social contract gives vastly better terms to other bargainers) (Cudd 9).
day. The billionaire makes it clear that actually drinking the stuff is irrelevant, he does not care whether you drink it, and will in fact deposit the money in your account at 12:01 A.M. once you make the sincere promise at midnight. Supposing that you are very unlikely to be able to fool the billionaire with an insincere promise (which corresponds to Gauthier’s assumption that straightforward maximizers cannot count on concealing their dispositions), it would clearly maximize your expected utility to form, if you could, the intention to drink the toxin tomorrow so that you can sincerely promise to do so at midnight. Suppose that you do so, and have collected the million at midnight. When noontime comes, is it rational to drink the toxin and make yourself very sick for a day?


Gauthier’s doctrine of constrained maximization would dictate that it is rational to drink poison with no possible benefit to you in doing so. At this point, Gauthier’s interpretation of Hobbes proves not only to have absurd consequences, but to also have completely misunderstood rationality as something with the potential to be knowingly irrational (as opposed to accepting limits of knowledge, and making rational gambles that may not turn out to be in one’s perfect self-interest, but were nonetheless rational to make given). However, Kavka claims that this problem is easy to solve from the vantage point
of his own, rule-egoist Hobbesian theory, which would simply deny that swallowing this poison would be irrational in this instance where one has perfect knowledge of the consequences of his actions (namely, nothing negative).

Nonetheless, Kavka’s own toxin puzzle may pose a more fundamental challenge of Hobbesian theory, including his own variant thereof. If it is truly rational to violate one’s rational agreements when one is assured of such a violation’s benefit, we have only escaped the normativity objection insofar as there is uncertainty as to the success of one’s law breaking. There is nothing, however, inherently wrong about successful criminal activity, which runs contrary to the general ideas government holds about law and justice. Thereby, the Normativity objection still holds. Kavka just denies that it’s as devastating an objection as one might suppose. Nonetheless, justice would no longer follow Hobbes’ definition as explored in part I. Rather than justice being defined as obedience to law, it would be defined as obedience to law in all situations where one would not truly derive greater benefit from breaking it. Laws are only “laws” in the weak sense that they are generally best to follow, not in the sense that they are mandatory.

If neither Gauthier’s constrained maximization nor Kavka’s rule-egoism can be proved to be directly rational to an agent, nothing binds agents to any rule that would lead them to constrain their self-interest (Cudd 2). To attempt to account for loopholes in obedience and the weakness of laws in such accounts, some neo-Hobbesians have taken the approach of agreeing with the Foole himself. This is a tactic articulated by Jean Hampton, though it is important to note that Hampton herself never claimed to be a neo-Hobbesian, or asserted the correctness of Hobbesian contractarianism, despite having articulated a revised version of his theory that she believed was the most generous
interpretation of Hobbes possible (Kraus 1993, 41). According to Hampton, expected utility maximization is the ruling motivation of Hobbes’ agents, who cannot break away from this motive once within the social contract (Hampton 1986, 16). Referencing Hobbes’ claim that some rights, such as immediate defense from harm, cannot be laid down because it would never be in one’s rational self-interest to do so, Hampton states “according to Hobbes, contractual obligations exist only insofar as it is in our interest to perform [them].” This implies that once these obligations cease to be individually rational, they cease to be binding according to justice (56). Rather than denying the substance of the normativity objection, Hampton’s Hobbes merely clarifies its implications (57).

In this way, while it is possible that neo-Hobbesian can narrow the scope of concern presented by the normativity objection, they cannot deny its claim. If successful breaking of the law is permissible, what is the substance of a state or social contracts? A Hobbesian society might have some of the same outcomes of behavior as a society based on less self-interested premises, but would ultimately find itself unable to demand obedience where disobedience can be successfully, and advantageously committed. At most, Hobbesian contractarianism can give individuals reasons to dispose themselves to act and/or reasons to appear to act in accordance with the law. However, given problems of collective action, such as the previously mentioned free-rider problem, this would lead to an unstable, provisionary system of governance (Southwood 2010, 35). This problem stems from Hobbes’ attachment to rational self-interest as the singular political and ethical justification for action and for government (Kraus 1993, 42), a type of ego
monism that certain Kantian contractarian theories, such as those of John Rawls, do not share.
Part III

Power, Inability, & Proximity
One of the main attractions of social contract theory has to do with its perceived inclusivity and commitment to a certain measure of equality among all subjects to a contract. Neither Kantian nor Hobbesian camps of social contract theory have an exclusive claim to this premise of equality, which exists in both camps and in social contract theory outside of this divide. While technically unaligned with either of the classical factions of contractarianism, John Locke most famously advocated a type of moral equality within the social contract that would lead to a sort of impartiality of the resultant laws.

Man being born, as has been proved, with a title to perfect freedom, and an uncontrouled enjoyment of all the rights and privileges of the law of nature, equally with any other man, or number of men in the world, hath by nature a power, not only to preserve his property, that is, his life, liberty and estate, against the injuries and attempts of other men . . . . And thus all private judgment of every particular member being excluded, the community comes to be umpire, by settled standing rules, indifferent, and the same to all parties.

-- John Locke, Second Treatise of Government, Ch. IV, Sec. 87

---

10 Though, traditionally defined as only white men.
While Kantian contractarians assert a similar moral equality of man and resulting “indifference” of law, moral equality is not the only way in which social contract theories can lay claim to “egalitarianism.” As discussed in Part I, Hobbesian social contract theories assert rough equality of capacity, and for this reason are presumed to be inclusive of all governed agents in the contracting position. Rather than claiming that all people are morally entitled to being included in a social contract theory, Hobbesians assert that it is in everyone’s rational self-interest to be inclusive, so as to negate the potential threats of any non-contractors who could otherwise, in the state of nature, inflict harm and act against other individuals’ self-interest.

Though there bee found one man sometimes manifestly stronger in body, or of quicker mind then another; yet when all is reckoned together, the difference between man, and man, is not so considerable, as that one man can thereupon claim to himselfe any benefit, to which another may not pretend, as well as he. For as to the strength of body, the weakest has strength enough to kill the strongest, either by secret machination, or by confederacy with others, that are in the same danger with himselfe.

– Thomas Hobbes, *Leviathan*, I.XIII.i

Connected to this commitment to equality, of whatever nature it may be, is a further commitment to the consent of each governed person. No matter the variant of
social contract theory, the need for the consent of the governed is always seen as a basic necessary condition for any social contract. As discussed in Part I, the definition of “consent” is highly variable from theory to theory, but its generality and scope remains the same, namely, that consent must be had from all. As Jeremy Waldron asserts, the unifying feature of the very diverse, and often internally antagonistic, tradition of “liberal” social contract theory can be understood as, “a conception of freedom and respect for the capacities and the agency of individual men and women...these commitments generate a requirement that all aspects of the social should either be made acceptable or be capable of being made acceptable to every last individual” (Waldron 1987, 128).

However, these assertions of equality, whether of capacity or moral entitlement, even with the extra provision of required consent, are not always enough to require inclusion of all governed parties in the contracting process. In many cases, the role of self-interest in the grounding of social contract theories, even with the above described constraints, would seem to permit if not require the exclusion of certain disadvantaged parties for the benefit of those included. Rather than finding legitimacy in the consent and equality of all parties, some social contracts would support themselves upon a foundation of exploitation, justified by the advantage to be gained by the unexploited.

While Hobbesian, Kantians, and others endorse some sort of equality of man, and thereby require the consent of each subject to a social contract, it doesn’t follow that all persons that are subject to a society and its rules must be treated in the exact same fashion in all cases, but rather that all people subject to a society should be treated in accordance with some sort of political principle to which they would agree. Moreover, while the principles of consent and equality of man extend to all persons within the social contract,  

11 In later theories and revisionist accounts of older doctrines
these same principles can make social contract theories quite exclusive by limiting those to whom they offer membership.

This leads to a general line of argument against contractarianism sometimes referred to as the “impartiality objection” (Southwood 2010, 40), which varies in nature depending on which parties one is concerned about excluding. In this section, I will address three different versions of the impartiality objection: one concerning relative power and unequal advantage, the second having to do with the inability of some parties to participate in contracting (hypothetical or actual), and finally the limits of scope of a contract based on region.
Chapter I: Power and Advantage

It is not clear how a social contract based on mutual advantage can account for the rights of those who do not provide any mutual advantage to other contractors, or those whom the other contractors would be more greatly advantaged by excluding from the social contract. Those with much greater power, say, a powerful majority, might find their individual self-interest better served by not contracting with a group with lesser power, such as a societal minority. At best, a contract based purely on self-interest without any mitigating considerations can only provide for the rights and well being of the “truly powerless” person if it is in the interest of the “truly powerful” (Southwood 44). This makes any concept of “rights” within a political system for those whose inclusion is not advantageous to others a superficial one, because it would be contingent on and derivative of the interests of the more powerful contractors (50).

Under Hobbesian contractarianism, to include all people within a certain geographic space in a social contract, we must assume that all parties are equally threatening enough that contracting with them is in everyone’s self interest. Hobbes assumes that this is, in fact, the case. According to his theory of man, humans are relentlessly in pursuit of power and so will take advantage of opportunities to harm one another so as to secure that self-interest (Hobbes 42). In addition, while men have unequal talents, none are so powerful that a less powerful person couldn’t do them harm by some means in a moment of vulnerability (54). Since one of men’s primary motivations in life is fear and the avoidance of potential harm (44), everyone will see their self-interest is best served by contracting with everyone else.
However, is this a realistic way of looking at disparities of power? Some critics, such as Nicholas Southwood, deny such a claim outright:

[T]he actual world is rife with discrepancies of power, strength, dominance, wealth, and so on that are sufficiently large to make it the case that there are some individuals (call them ‘the Truly Powerful’) and some other individuals (call them ‘the Truly Powerless) such that it may not be in the interests of the Truly Powerful to agree to constrain their conduct towards the Truly Powerless.

-- Nicholas Southwood, Contractualism and the Foundations of Morality, 45

It’s not hard to see the force of this characterization of unequal power. Consider children, the mentally or physically disabled, the agoraphobic, etc. We can certainly imagine cases in which a person would, voluntarily or involuntarily, fail to pose enough of a threat that it would be in our self-interest to neutralize them with a social contract.

One possible way of dealing with this concern is simply to say that we will not contract with those people who are not threatening enough to be worth contracting with. However, this allows us to treat all those left outside the contract without any sort of considerations of justice, since, according to a Hobbesian system, right and wrong only exist within a social contract, which is why we so desperately need to form one (Hobbes 2012 [1651], 19). For the truly powerless, the conditions have not changed: they are not
in the contract and they remain in the state of nature. Therefore, we are not bound to respect these persons or their interests and can do anything to them that is in our self-interest. This leads to a conclusion that is not easy to accept, and does not offer the advantages of rough equality and respect for consent that normally serve as the attractive aspects of social contract theory: that there are “some individuals who are de facto entitled to do anything whatsoever to some other individuals” (Southwood 2010, 45).

There have been some suggestions as to why, even in the case of persons who are unthreatening (and don’t provide other advantages as contractors), Hobbesian or Hobbesian-inspired contractarian theories would still require inclusion. A common idea is to add to Hobbesian contractarianism a Lockean proviso. While John Locke in his contractarian theory also believes that people have a certain right to pursue their own self-interest (in his case, he focused on the acquisition of private property), he described a certain limit on this self-interested pursuit. Namely, that a person might pursue their ends (or acquire private property) so long as it does not entirely deprive others.

“So that, in effect, there was never the less left for others because of his enclosure for himself. For he that leaves as much as another can make use of does as good as take nothing at all. Nobody could think himself injured by the drinking of another man, though he took a good draught, who had a whole river of the same water left him to quench his thirst. And the case of land and water, where there is enough of both, is perfectly the same.”
While John Locke himself never elaborated on such a “proviso,” a term first used by Robert Nozick in his book, *Anarchy, State & Utopia*, it has become a common idea in social contract theory (Nozick 175). This proviso is used to describe the procedure by which unowned things become owned, a process known as “initial acquisition.” For the case of Hobbesian contractarianism, we might argue that if resources were acquired unfairly or sufficiently unevenly to begin with, the unequal standing of those in the contractarian position would be unjust, and therefore in order to be just the social contract must be more inclusive than a simple calculation of rational self-interest at the hypothetical moment at which one would form a social contract to escape the state of nature (Southwood 2010, 44).

It isn’t open to Hobbes himself to employ a Lockean proviso, given his insistence that justice is exclusively based on self-interest. It is unclear why rational self-interested agents, in mortal fear of each others’ destructive powers, would consider it worth their while to leave “as much and as good” resources for weaker individuals. Moreover, as Nicholas Southwood\(^\text{12}\) argues, it is unclear that a Lockean proviso would do much to solve the problem at issue in Hobbesian contractarianism. In fact, Southwood claims that such an addition would be both too bold and too weak to defend this contractarianism from the impartiality objection (Southwood 47). To assert that a Hobbesian contractor

\(^\text{12}\) It is important to note that Southwood is discussing is a type of moral contractarianism, rather than a political contractarianism. I am appropriating his arguments to see how they might apply to social contract theory. I make no stance in this thesis as to whether political philosophy is, in fact, a branch of moral philosophy, or about the worth and correctness of moral Contractarianism.
should respect a Lockean proviso, even when it is not in their rational self-interest, contradicts Hobbesian contractarianism’s explicit and fundamental assumption that rational self-interest as the singular reason for contracting in the first place. Conversely, even if a Hobbesian contractor were to adopt such a proviso, that would not constrain them greatly from excluding those who it is not in their self-interest to include. For example, inequalities due to unequal resources, so long as the advantaged party did not do anything to prevent others from using the resources available to them\textsuperscript{13} would be allowed. Inequalities due to differences in talent or luck would not be ruled out as a legitimate basis for exclusion from a place in the contracting position, potentially leaving some able to either exclude others or create a contract that did little to protect them (Southwood 2010, 45).

Hoping to escape the objectionable outcomes and implications of its Hobbesian counterpart, Kantian political contractarianism places a special emphasis on certain standards of equal treatment of persons. Possibly the best known is that of *Justice as Fairness*, the theory of justice created by John Rawls (described in part 1). As I explained earlier, Rawls argues that in order to form a social contract justly and correctly, those undertaking the project must be (imaginatively) extricated from their present, existing social structures (Rawls 542). Therefore, during the Rawlsian contracting process, which Rawls refers to as the Original Position, there is a certain hypothetical state of ignorance which Rawls did not believe was actually achievable, but which, if considered in the hypothetical, would still provide crucial insight into social justice (Rawls 544). Rawls’ ideal “veil of ignorance,” would make any knowledge of morally irrelevant personal

\textsuperscript{13} assuming different resources are generally available to different parties, due to proximity if nothing else.
characteristics and circumstances inaccessible to participants. By stripping these deciders of all morally irrelevant factors, such as their own class, abilities, education, religion, social connections, status, and personal conception of the good, Rawls intends to remove all personal biases and leave only knowledge of what is relevant to considerations of justice in general.

While Rawls himself did not specifically mention race, gender and sexual orientation in his accounts of morally arbitrary features that would be unknown behind a veil of ignorance, many of his interpreters and later Rawlsian scholars have taken this to be the case. For example, Thomas Pogge argues that such features, along with a myriad of other features that facilitate societal discrimination, are not mentioned simply because Rawls is engaged in the project of ideal theory (rather than a historical analysis of injustice), where race and gender create no claim to moral preference or disadvantage. In real (practical) theory, one might consider the existence of actual prejudice, but for the purposes of ideal theory, discrimination based on biased beliefs about genders, races, sexual preferences, and so forth, has no place. Furthermore, Pogge claims that such inequalities are naturally economically inefficient in maximally promoting the primary goods position of the worst off, so would be doubly prohibited in a Rawlsian system (Pogge, John Rawls: His Life and Theory of Justice 2007, 124).

However, even behind the veil of ignorance, relevant considerations do include certain kinds of knowledge of the world and of human nature. For example, the limited altruism of human beings and the moderate scarcity of resources would be facts about the world of which those behind the veil of ignorance would still be aware. The first of these is a large departure from Hobbesian and neo-Hobbesian contracts, as is the veil of
ignorance device in general. Hobbesian contracts allow for fear and intimidation due to an existing power struggle to serve as the motivation of the contract, during which each contractor is entirely aware of themselves and their specific, personal motivations. As explored in Part I, Rawlsian contracts break from Hobbesian contracts in two key ways:

1. They differ in the nature of the contracting situation. Hobbesian contracts presume a starting point of “the state of nature,” that includes knowledge of one’s position. In contrast, in a Rawlsian initial position, Rawls sets certain parameters of knowledge (the “veil of ignorance”) so as to prevent people from trying to form a social contract based around inequalities that were not themselves formed in a fair social contract, and would not necessarily be just enough to use as a foundation for a new social contract.

2. The motivations of the parties to the contract differ between Hobbesian and Rawlsian contracts. Rawlsian contractors not only are rational like their Hobbesian counterparts, but are also reasonable. For Hobbes, in contrast, the misery necessarily resulting from a world without law or sovereign will be all the motivation rational agents need to contract socially. The resultant contract, so long as it does not demand the impossible (such as that a man not defend himself when threatened with death) will be freely agreed to and will therefore be just.

Rawls sees ignorance of morally irrelevant characteristics as especially important to defining justice because these characteristics overwhelmingly determine the degree of
fairness and impartiality of a society. It is hoped that if all contractors are ignorant of their relevant power, an exploitative contract, such as those potentially arising from Hobbesian contractarianism, can be avoided. Unfair and arbitrary inequalities pervade the structures of societies, so they must be the first things addressed by a social contract that is formed for the purpose of promoting justice (543). The equality Rawls hopes to establish covers all potential inequality of power in the contracting position that posed problems for Hobbesian contractarianism by suspending not only inequality of physical resources, prejudices concerning characteristics (race, gender, sexual orientation), but even differences of talent, utility, or potential advantage to a society and other individual contractors.

Rawls finds such awareness counterproductive when attempting to form a just social contract, because the relative positions and needs of persons are morally arbitrary, and thereby cannot create a foundation for political justice. In the case of social inequalities, as Rawls describes it, “not only are they pervasive, but they affect men’s initial chances in life; yet they cannot possibly be justified by an appeal to the notions of merit or desert” (543). We may all agree that some characteristics of a person’s situation at birth, such as differences of race, gender, and sexual orientation, are not determinant of a person’s moral worth and ethically should confer no advantage. We will be led by logic to accept that any disadvantages or advantages produced because of those characteristics cannot be morally justified.

Some critics object that that natural abilities, education, natural gifts such as beauty, and other traits should be rewarded by society, because they have positive effects, or because those inequalities are inherently fair and just. There are two competing
understandings as to why Rawls treats talent in the manner that he does. However, both lead to the conclusion that talent is morally arbitrary in some sense. One interpretation of *Justice as Fairness*, sometimes referred to as “luck egalitarianism” or the “no control” argument, asserts that inequality based on natural gifts cannot be just simply because nobody *chooses* whether or not to be talented, beautiful, or otherwise gifted as a matter of birth. Rather, such features result from a natural lottery over which an agent does not have control. Rawls contends that if an agent has no control over his allotment of these characteristics, he cannot be said to be morally responsible for his flaws or morally deserving of his talents, as he did nothing to deserve or earn them (Pogge 74-75, 80-82, 116, 123). Again, these unearned advantages are the source of inequality and therefore of social injustice, and must be dealt with in a just society. On the other view, these natural talents are morally arbitrary not simply because they are not chosen, but because the mere existence of natural gifts does not in itself generate a claim to any particular advantage. For example, the mere ability or potential of someone to run faster than other people does not entitle them to an Olympic championship (Armstrong 14). Although this distinction is important for interpreting Rawls’ theory more generally, it is not pertinent to my discussion of the exclusionary objection, as both interpretations lead to the conclusion that inclusion in the contracting process, or one’s equality within that process, should not be determined by natural gifts. I will take no further stance here on the issue of which interpretation is correct.

It is important to note here that not all people feel that Rawls’ veil of ignorance provides sufficiently for the inclusion of all within the social contract, particularly those who are “non-productive” economically. As Martha Nussbaum states it,
“This emphasis is built deeply into the logic of the contract situation: the idea is that people will get together with others and contract for basic political principles only in certain circumstances, circumstances in which they can expect mutual benefit and in which all stand to gain from the cooperation. To include in the initial situation people who are unusually expensive or who can be expected to contribute far less than most to the well-being of the group (less than the amount defined by the idea of the "normal," whose use in Rawls we shall study shortly) would run contrary to the logic of the whole exercise.”

-- Martha Nussbaum, *Frontiers of Justice*, 104

Nussbaum contends that the very structure of social contract theory is so intertwined with amoral individual interest that it can never fully construct a just system of social order. She sees inequality of ability as one of the “unsolved problems of justice” which social contracts cannot answer.

In social contract theory’s defense, Rawlsian scholars, such as Samuel Freeman, show that such worries about Rawlsian theory not providing for the “less productive” actually derive from a conflation of Rawlsian theory with Hobbesian theory, rather than from features inherent in a Rawlsian social contract. As discussed in Part I, rational interest is a tool used within the limits of the original position with a veil of ignorance.
Therefore (leaving aside the inabilities of those without the mental capacity to rationalize within a social contracting position: the subject of the next chapter) disabilities can be excluded from parties’ knowledge, so that the needs and goals of disabled persons would be provided for as any demographic group within the original position. This contradicts Nussbaum’s view that Rawls sees the purpose of social cooperation as gaining mutual advantage, which would be more properly attributed to Hobbesian theories (Freeman December 2006 9).

Instead, Rawls criticizes and breaks with such a Hobbesian conception of social cooperation as justified by mutual advantage, citing the importance of human beings’ social nature, a sense of justice, and “reasonableness” as explored in Part I. There is no question for Rawls as to whether to include less economically productive parties in the social contract, and furthermore, relative productivity and offered advantage would be concealed behind the veil of ignorance in an effort to preclude discrimination or inequality based on those features during the contracting process. Justice as Fairness simply does not promote discrimination or exclusion from the social contract based upon such differences:

“Associations and individuals have such ends, but not a well-ordered society; although it has the aim of giving justice to all its citizens, this is not an aim that ranks their expected contributions and on that basis determines their social role or their worth from a social standpoint. The notion of an individual's contribution to society viewed as an association (so
that society is entitled to offer terms for joining derived from
the aims of those already members of the association) has no
place in a Kantian view. It is necessary, therefore, to construe
the social contract in a special way that distinguishes it from
other agreements.”


Despite this clarification, the needs of severely disabled persons will still pose
substantial problems for Rawlsian theory, which will now be explored in the following
chapter.
Chapter II: Inability

Even with the issues of unequal power addressed, there is still a glaring instance of exclusion in both Hobbesian and Kantian formulations of social contract theory. Not only should we be concerned with whether all morally relevant beings would be advantageous to contract with, but also with whether they would be capable of contracting and demanding that their self-interest be considered. Even if those behind the veil of ignorance cannot screen out the self-interest of those with less societal power than them, what prevents them from neglecting those who, because of sheer inability or disability, cannot take part in a rational contracting process? Even if those in the contracting position found it in their self-interest to provide for those unable to contract, wouldn’t the rights of those parties derivative, making people morally unequal within a social contract, and thereby within the resultant society?

It is important to note here that the only people that could potentially be subject to the inability objection would be those in which some persons had severe mental disabilities. Mere differences in talent or intelligence are more appropriately dealt with in the previous chapter, as a problem of the difference of relative power. Severe physical disabilities, so long as they do not make a person rationally and mentally incapable of contracting, would not prevent contracting in the same way. While Hobbesian contract might find such persons with physical disabilities to be so unthreatening that they are not rationally worth contracting with, these persons would not be themselves incapable of contracting. Hobbes is very clear that while there are differences in ability between people, so long as they are even somewhat dangerous to other contractors they are still
party to the contract (Hobbes 55). While this formulation of the issue still leaves many in an unequal bargaining position, it does so not because those persons are less able to contract, but because it is not in the particular self-interest of the more powerful parties to give equal bargaining status to those persons. Thereby, differences in intelligence and talent, while they do pose a problem for Hobbesian contractarians, are not problems of inability but of unequal mutual advantage.

As discussed, Hobbesian contractarianism finds itself only somewhat capable to provide for those with fewer natural talents, and outright admits that exclusion is an acceptable option where it is rational. This might make the rights of the disabled contingent or derivative of the self-interest of other parties. According to David Gauthier, Hobbesian contractarians may simply have to assert that disabled persons, along with anyone who is not rationally advantageous to contract with, are simply not entitled to the relationships created in a contractarian theory. Any sort of remedial justice or aid to disabled persons is viewed as a drain upon the rational self-interest of contractors, who are not in any way obligated by self interest to agree to principles that would coerce them to uphold any sort of rights or standard of living for disabled persons (Gauthier, Morals by Agreement, 18). As exclusion is already admitted as a politically acceptable option for a socially contracted society, this detracts little from Hobbesianism that has not already been relinquished.

Two types of criticisms arise when one addresses the place of disabled persons. First, and most relevant in the Hobbesian case, is that the rights and standard of living of the disabled would be, at best, derivative of other persons’ rational self-interest and

---

14 In Gauthier's case, he was exploring a moral theory, but the same blunt denial of contractarians' rights for the disabled is applicable to a Hobbesian theory.
therefore insecure, rather than being something inviolable and universal. In addition, we might charge social contract theory of making the rights and needs of the disabled a morally secondary (i.e., less important) concern, separate and different from those of other human beings (Nussbaum 2007, 16). This is a distinctly different objection, occasionally used against Kantian theories, which I will treat later in this section.

While the difference between an inability objection and an exclusionary objection based on relative power may prove negligible for Hobbesian theories (as they readily accept some exclusion) the distinction proves dramatic for Kantian theories. While Kantian theories, as explored above, don’t have as much difficulty accounting for a lack of self-interested motivation to contract, a rational inability to contract poses different sorts of challenges. Admittedly, in Rawls’ view, the mere existence of unequal natural assets, including that of intelligence, is not a sufficient justification for exclusion from a social contract. However, Rawls does make some assumptions about the character of the contractors. Most significantly here, the ability to reason is necessary to contract in the first place (Nussbaum 2007, 54).

Reasoning to promote one’s self interest behind a veil of ignorance is not a feat all parties could take part in, and therefore there are non-contractors affected by the social contract. The severely mentally disabled for example could not take up such an impartial, rational position, and thereby wouldn’t be considered as one of the potential persons one could be once the veil of ignorance is lifted. Thereby, it is not necessarily in the self-interest of any of the impartial contractors to form the contract in such a way that provides sufficiently for those with such impairments. This problem is faced not only by the mentally impaired, but also by any non-contractor, including non-humans and persons
outside the nation that is formed by the social contract in question. According to social contract theory, we are only obliged to contract with those who can demand in a contracting position that their mutual self-interest should be taken into account, and will only provide for the self-interest of others if they are party to the contract. Aside from this, social contract theory does not define or limit how one treats other human beings, not part of the contracting process. Can the parties within the contract, if they are to be disinterested in other people’s interests, provide for those that are non-contractors? Furthermore, should they?

It does not initially seem so. Rawls is rather clear that charitable impulses can play no part in the motivation of the contractor in the original position: “The parties do not seek to confer benefits or to impose injuries on one another; they are not moved by affection or rancor” (Rawls, A Theory of Justice: Revised Edition 1999, 124). Therefore, what falls outside a consideration of self-interest also falls outside of a contractor’s consideration. Agents are not to take any interest in the interests of others – which would preclude them from considering the needs of those they could not potentially find themselves in the position of when they come out of the contracting position. Even if such groups, outside of the social contract, did provide some mutual advantage, they would only be considered as having rights to the extent that the contractors were benefited by those rights, which would make the standing of disabled persons, animals, and citizens of other nations not intrinsically deserving of consideration, but rather contingent on the self-interested of a privileged group able to contract within the original position.
To focus on the first of these troubling cases, in no social contract theory are those persons with severe mental disabilities included among those contracting. In fact, there couldn’t be a way to include them given the structure of contractarianism: “*Some people with severe mental impairments, however, could not be included in the group of political choosers directly, however generously we assess their potential for such a contribution*” (Chapter I, section ii, part 1, FOJ Nussbaum). There is no good reason to believe that such people would be provided for despite the lack of such contractual consideration, as evidenced by societies’ consistent failure, up until recently, to do so. Even if, through charitable impulses outside of what are demanded by the fundamental social contract, people were to provide for these groups, their care would be contingent on “true” contractors, and would not be, in any meaningful sense, “rights.” Their protections under the social contract would be derivative and less fundamental than considerations for typically abled persons (Nussbaum 16).

According to Nussbaum, there is no way for even the most refined social contract theory to overcome this objection: “*When we discuss mental disability, we will see that the equation of a citizen status with (prudential and moral) rationality is a hurdle that even the best contemporary theories cannot surmount, without losing their formative link to the social contract tradition*” (Nussbaum 2007, 54). In her view, there is no way for contractarianism, Rawlsian or otherwise, to include the rights and needs of the severely mentally disabled equally in the social contracting position. Even if, somehow, they were able to be included, the principles Rawls asserts would not cater to their needs. For example, Rawls’ difference principle, with its emphasis on the distribution of primary goods, and the presumption of equality of these goods except in situations where it
benefits the least economically advantaged does not account for disadvantages other than lesser amounts of primary goods (specifically, the disadvantage of disability) nor does it acknowledge that differing amounts of need between persons (acknowledging that those with disabilities with require greater amounts of primary goods to achieve a similar standard of living to other citizens). Income and wealth, which are the primary if not exclusive focus of the Difference Principle according to Nussbaum, are simply not sufficient to gauge the needs of the “worst-off” in terms of disability, and will lead to insufficient concessions or considerations for them (Nussbaum 2007, 115).

There are several ways to defend Rawls from these two lines of critique. One way is to have those within the social contracting position consider the creation of a hypothetical insurance policy for those with severe disability, not being aware of how many disabled persons each contractor represents (so, having no reason to withhold benefits for their own advantage) (Freeman 2006 19). In the case of disabled persons, it is sufficient for contractors to be aware that they themselves could one day find themselves in the position of a disabled person. As they are aware of the basic facts of life, including the tendency of human beings to humanly identify with their children, they would be concerned also about the prospect of having a child with such a disability, and would want to secure their rights. Finally, they will be aware that a good number of contractors, including possibly themselves, will have benevolent feelings concerning the disabled in a well-ordered society, and will rationally choose to support their needs and liberties. Thereby, in several ways, contractors could recognize the humanity and moral worth of disabled persons through identification and the potential for their own personal interest in such persons’ welfare. This would require some stretch of the imagination, but nothing
more dramatic than the stretch of the imagination that Rawls’ original position already demands (19).

As such contractors would be representing those not in the contracting position, this stance is necessarily connected to that of trusteeship for severely disabled persons. However, an argument from trusteeship, though it makes some concessions to the needs of the disabled, might still be subject to the objection that those needs would still be derivative of the choices or needs of those included directly in the social contract. If those within the social contracting position are as truly self-interested as Nussbaum believes, there might not be a way to convince those parties to treat those they represent with equal consideration to themselves. While an assertion of trusteeship would not be impossible within a Rawlsian view, there is reason to think that Rawls had a different solution in mind when considering the needs and rights of severely disabled persons (21).

Such an alternate argument, which does not appear to be compatible with the idea of trusteeship, is to create separate realms of justice, of which the sort of distributive justice dealt with by the difference principle is only one. The needs of the disabled, as well as many other types of concerns of justice, would be more properly dealt with in other stages of the social contracting process. According to Rawls, the original position and two principles of justice created behind it are made to address certain “background” principles of justice for the “basic” structure of society, which include the distribution of resources among rationally capable and economically productive people. While the two principles of justice don’t address severely disabled persons, a certain measure of remedial justice for such groups are dealt with at later legislative stages of the social contracting process.
According to Freeman, Rawls’ definition does not include any persons who are not economically productive. People who are homeless, unemployed, too disabled to work, etc. are not present in Rawls’ exploration of distributive justice. Instead, Rawls opts to deal with these parties at a different, though not less important, part of the contracting process, making these considerations of “remedial justice” rather than distributive justice (Freeman, Rawls 2007, 106). In effect, this would make the contractarian process in Rawlsian original position much less comprehensive than other social contract theories, such as those of Hobbesians, which intend to justify all law and order based on a contract of rational agents.

This is not to imply that the issues concerning the severely disabled are less important or a secondary consideration of justice, but that they are an issue of justice that is best dealt with separately: “These basic social institutions are necessary for a complex society to function, adapt to changing circumstances, reproduce itself, and endure from one generation to the next. In this regard, remedial institutions and agencies for the disabled, like other important social institutions, such as religion--however profound their influence on people's lives--are not basic institutions” (Freeman, 2006 20). In fact, there are many kinds of justice, admittedly political, that Rawls did not assert that his social contract could solve. For example, in the Law of Peoples, Rawls not only asserts that there is a separate remedial justice that we have duties to enforce, but also a natural justice to support people’s human rights, including minimum economic security and basic liberties, simply because of our shared humanity (Freeman, Rawls 2007 107-108).

There might still be some concern that treating disability justice separately from “basic justice” invites discrimination and encourages the view that the concerns of the
disabled are either secondary and/or derivative. However, there seems good reason to
treat justice to those so cognitively impaired so as to not be able to reason in a collective
in a different way than we treat justice towards those with such rational capabilities. One
might attempt to defend Rawls by arguing that it is only just that these persons are treated
differently by a society, as they are not held to that society’s usual standards. In the case
of the former, severely mentally disabled persons are able to evade certain
responsibilities of citizenship, such as certain punishments for actions that, if committed
by someone without any such disability, would be treated as a crime. This view is
compatible with what is previously described, as one can sympathetically identify with
severely mentally disabled persons because it is a potential outcome of one’s own life.
However, the contractors can reasonably limit the liberties and responsibilities of these
persons given their capabilities and still provide for a certain measure of liberty and
wellbeing of such parties by taking into account the difference in how such persons
interact in a society: “It is not normally an injustice to deny the severely mentally
disabled such basic liberties as unrestricted freedom of movement, freedom of
association with whomever they please, the right to enter into contracts, and equal
political liberties, for these rights are exercisable only if a person has developed moral
powers and is capable of understanding his or her own interests. But it is a gross
injustice to deny the mentally disabled their human rights” (Freeman 2006 18).

While a Rawlsian version of Kantian contractarianism is in this way able to
include the needs and rights of the severely mentally disabled in a way that does not
imply that their rights are inherently less important than those of others, in doing so it
relies on the assertion that there are human rights and liberties that are owed to all human
beings, rather than being derived from a contract between each member of a society. This limits the scope of what contractarianism can account for as a political theory, making it less than comprehensive. On the other hand, Hobbesianism merely rejects the need for such inclusion, allowing itself to only be “comprehensive” in the sense that it denies all duties of justice for which it cannot adequately account. In this way, Hobbesian contractarians beg the question of their own correctness about a sense of justice, stating that there can be no justice to the disabled simply because they cannot account for it. While the former, Rawlsian theory, is in this way preferable to Hobbesian contractarianism from the perspective of political justice, its ability to be inclusive comes at the price of limiting the role of contractarianism in general.
Chapter III: Outside the Contract

“However intimate the connection may have been between justice and the state in the world that Hobbes (as well as Rousseau, Hegel, Mill, and Morgenthau) occupied, and whatever we may think of the victory of modern accounts of sovereignty and justice over a tradition of “associative justice,” which rooted norms in a variety of forms of human association not confined to the state, it is now a mistake to assign the state so fundamental a role in political morality.”

– Sabel & Cohen, Extra Rempublicam Nulla Justitia, 149

It is commonly assumed that justice doesn’t end at the border. However, such a seemingly common-sense understanding of justice may collide with the very contractarian foundations upon which our sense of justice is based. According to contractarianism, we base our political duties on the consent of the governed, of our co-contractors, and use this as a comprehensive justification for supporting a governmental structure and our sense of political justice. However, we tend to then extrapolate our ideas of political justice, and evaluate situations in other nations based on this sense of justice. We take ourselves to have some political responsibility towards members of other nations (refugees, for example, and possibly those being persecuted by their own governments), which is not entirely dissimilar to the responsibility we feel for our own citizens. Furthermore, we often justify our duties towards these people with our political
principles (rights to life, liberty, etc.), even though those persons are not technically within our national social contract. How can we account for this political action and responsibility on the part of those outside of the social contract?

Persons who are not citizens of the society being contracted are by definition outside of the social contract (Chapter I, section ii, part 1, FOJ Nussbaum). The place of the members of other nations in the social contract touches not only on issues of impartiality, but also on those of international justice. The argument here is similar in nature to the other kind of impartiality objections explored in this section of the thesis. Those persons other than the contractors, those agents conceived as in the contractarian position, are only considered secondarily, derivatively, and/or out of virtue of charity rather than the same robust sense of justice accounted for in the social contract. However, there is one crucial difference: for many it doesn’t appear as offensive to our ideas of justice for those in a society to form such a contract, to have nations, and to assert “special duties” to their fellow countrymen in the same way that excluding disabled or less powerful persons seems obviously unjust.

For Hobbesian contractarians, this difference in our reactions is an appropriate response to the fact that justice simply does not exist outside of nations. If there is no sort of coercive sovereign unifying a people, people have no duties of justice whatsoever and remain within the state of nature. And there is no such all-encompassing sovereign in global politics. While Hobbes himself said little about the interaction of nations, when describing the power of the Sovereign, he wrote of international relations as something to be done solely for the benefit of the nation, rather than as requiring a mutual sense of justice. Hobbes said little more than, “Judging when it is for the publique good, and how
great forces are to be assembled, armed, and payd for that end [is the duty and right of the Sovereign]” (Hobbes 86).

When Hobbes speaks here of “the publique good,” he does not mean this to include the good of other nations. Instead, he implies that a sovereign ought to conduct international relations with a sole regard to the interests of his own nation. This has been developed into a view called Realism, which asserts that because there is no mutual governance of nations by a greater sovereign force, and because there is no social compact between them, there can be no international obligations of justice. This means that any sort of action a nation finds advantageous is permissible: States, “for their own security, [nations may] enlarge their dominions upon all pretences of danger and fear of invasion or assistance that may be given to invaders, [and] endeavour as much as they can, to subdue and weaken their neighbors” (XIX 4).

There are variations of this argument, usually surrounding the difference between it being permissible and it being mandatory to conduct international affairs exclusively according to national interest (Buchanan, Beyond the National Interest n.d., 222). However, for our purposes, it will be sufficient to focus on the idea that no obligations of justice exist outside of a social contract, as this is most intimately intertwined with the idea of self-interest as the basis of political, contractarian justice. Whether we are obligated or are merely permitted to act solely in our self-interest when making international policy, this Realist view will be dependent on the idea that we have no obligation to consider the interests of other nations, insofar as the interests of other nations don’t benefit our own national self-interest.

A basic Hobbesian realist argument can be formulated something like this:
Premise 1: No justice precedes the social contract and Sovereign enforcer.
Premise 2: Without justice, it is permissible to act solely according to self-interest.
Premise 3: There is no social contract with a sovereign enforcer over nations.
Conclusion 1: Justice does not exist between nations.
Conclusion 2: It is permissible for nations to act only according to their self-interest.

One might object here that there is no necessary equation between persons and nations. However, the extension of premise two to nations appears more natural when one considers that, according to Hobbes, human beings enter a social contract for the sake of their self-interest, a social contract solely exists only because of the self-interest of its contractors. This would lead the “nature” of these nations to be guided by self-interested motivations. As such, states are either permitted or obligated to act only with regard to their own citizens, and justice only exists between citizens, who must uphold the laws to which they have contracted. So-called neorealists such as Hans Morgenthau remain ideologically very close to this Hobbesian model. Morgenthau in particular emphasizes the power struggle in international relations, defining national self-interest as power and power alone (Morgenthau 1954, 5-12). Therefore, a good theory of international relations is one that is rationally guided by the goal of securing as much power as possible, with no thought to other goals such as peace, equality, non-violence, etc. (Korab-Karpowicz 2013).

One attempt to undermine this argument is to deny premise 3. As Thomas Pogge
argues, the severe inequalities between nations are actually a result of global social orders that, while they might not have the comprehensive power of a sovereign, have great influence over the organization of international relations. Such a global order has established hierarchies in the past and continues to enforce inequalities in the present. He cites border protection policies, economic and environmental externalized costs, unequal resource depletion and benefit thereof, international resource privilege\textsuperscript{15}, international borrowing privilege, treaties, alliances, etc. and international intellectual property rights as imposed social regulations that push certain nations below the “state-of-nature baseline”:

\begin{quote}
“However one may want to imagine a state of nature among human beings on this planet, one could not realistically conceive it as involving suffering and early deaths on the scale we are witnessing today. Only a thoroughly organized state of civilization can produce such horrendous misery and sustain an enduring poverty death toll of 18 million annually”

– Thomas Pogge, \textit{World Poverty and Human Rights}, 3
\end{quote}

However, the fact that international relations have shaped the structure of the world is not enough to show that the key features of a Sovereign – namely, punishment for transgressions and enforcement of some sort of law – exist or are emerging globally. Nonetheless, some find that emerging forms of international governance are increasingly

\textsuperscript{15} Define terms here (2x)
playing such a role. While states are assumed to be largely mutually isolated in many political contractarian theories, the coercive powers of a sovereign are ever more present in international rule-making bodies. Joshua Cohen and Charles Sabel, in their refutation of neorealism, summarize this view of current global politics:

“While states remain essential players, to a considerable and growing extent, rule making, as well as rule elaboration and application, especially in the arena of economic regulation, but also in areas of security, labor standards, environment, rights, food safety standards, product standards among others, are taking place in global settings that, even if established by states (and many regulatory functions are provided by private or public–private bodies), conduct their activities of making, elaborating, and applying rules activities with some de facto decision making independence from their creators.”

-- Cohen and Sabel, Extra Rempublicam Nulla Justitia, 165

Despite some current limitations, such as the lack of ability to directly and violently impose sanctions and law, the costliness of defying or withdrawing from supranational rule-making organizations is costly enough to be coercive enough to play the role of a sovereign.
While this argument and ever-increasing globalization undermine the empirical premises of Hobbesian Realism, the more fundamental argument against realism comes from a challenge to the basic premise that there is no justice outside of social contracts. Whether or not there is a global order, how could it be permissible to exploit and do any sorts of harm we wanted to people of other nations? Isn’t there some fundamental duty we have either not to harm them or to help them? In a Hobbesian system, there is simply no solution to this problem. As explored in part one, it is a fundamental premise for Hobbesian contractarianism that political justice exists only within a social contract created out of the self-interest of the contractors. However, more modern versions of a contractarian international Realism are willing to break with Hobbes enough to provide some justice for those outside the social contract, even if there is no world order.

One of the most recent and most controversial arguments for a limited Realism, concerning only distributive justice, comes from Thomas Nagel, who argues that the coercive nature and citizen self-authorship of state institutions are the only legitimate foundations upon which distributive justice is based. His position calls for a very strong break between the domestic realm, within which co-citizens owe each other a fairly substantial degree of equality because of their shared social contract and legal order, and the global realm, where our only duties to outsiders are minimal and humanitarian in form, not duties of socioeconomic justice. We may also have obligations of justice that concern things other than redistribution: “In a broad sense of the term, the international requirements of justice include standards governing the justification and conduct of war and standards that define the most basic human rights” (Nagel 114). Therefore, Nagel

---

16 In the sense that this objection concerns real theory, rather than the characterization of current global politics.
doesn’t reject global justice in its entirety the way a strict Hobbesian realist would, but merely rejects global distributive justice. This may be as close to a Hobbesian contractarianism that allows for some measure of global justice as possible, and even this relies upon breaking with certain fundamental tenants of neorealism and Hobbesian contractarianism.

On the other hand, Kantian theories are more open to the expansion of duties of justice to include people of other states. Charles R. Beitz attempts to globalize Rawls’ difference principle by excluding knowledge of one’s nationality behind the veil of ignorance when forming a social contract (Beitz 1975, 367). This would institute an international vision of contractarianism, based on the assertion that geographic distribution and access to resources is morally arbitrary (369), and that, so long as a global scheme of cooperation and interdependence exists, contractarianism should not consider borders to be of fundamental moral significance (376). This is compatible with Rawls’ idea of creating a second social contract. Coupled with Rawls’ assertion of certain existent natural and remedial duties owed to all human beings, there seems to be sufficient ground from a Kantian-Rawlsian perspective to assert that justice meaningfully applies internationally (Freeman, Book Review: Frontiers of Justice: The Capabilities Approach vs. Contractarianism 2006, 37).

If such independent duties of justice and political morality do exist outside of a contractarian position, as Nagel, Rawls, Freeman and other political contractarians wish to assert (to varying degrees), this in itself admits a certain limit to what social contract theory can provide in the establishment of a political and social order. While some duties may be able to be derived without any sort of reference to freestanding moral principles,
if we are to employ a general humanitarian responsibility to account for situations in which certain parties are unable to be adequately provided for by a social contract, we must admit that contractarian theories only apply within a certain scope, and are not comprehensive political theories. The assertion that there are basic moral duties, whatever they may be, towards all people is sufficient to effectively counter the impartiality and outsider problems. However, it requires admitting that social contract theory, even when it limits the role rational self-interest can play in founding justice, only applies to a limited scope of political interaction, and cannot be perceived as a comprehensive theory of political philosophy. This does not make the assertions of certain political contractarianisms outright wrong, but shows them to be incomplete.

**Convert all to either “Contractarian or Contractarian”**
Conclusion

In this thesis, I have examined the philosophical merits and political consequences of a self-interested social contract theory. As political contractarianism is by no means a uniform tradition, and has many internal disagreements, I have been selective. I chose to review the theory that gives the largest role to self-interest (that of Hobbesian and neo-Hobbesian political contractarianism) as well as a theory that only gives it a limited foundational role (Kantian political contractarianism as expressed by Rawlsians). After reviewing the essential differences between the two schools of thought in Part I, I moved on to one of the more fundamental problems with using self-interest as a foundation of political justice: self-interest does not in itself, oblige individuals to always act according to their agreements. If we allow not only that justice is supported and defined by self-interest, but also that violations of agreements made in the name of self-interested justice are supported in the same way, we end up with a strange and unstable understanding of justice and law. This objection only applies to Hobbesian versions of political contractarianism.

However, Hobbesian and Kantian-Rawlsian contractarianism both find themselves unable to guarantee justice for certain parties. In this thesis I specifically examined groups with less bargaining power, those persons so severely mentally disabled that they could not engage in self-interested bargaining, and those outside of the social contract. This is not an exhaustive list of groups that are either excluded or treated unequally in social contract theory. Importantly, there are also concerns about justice towards animals and the environment that are not easily resolved by the veil of ignorance.
or any of the other potential avenues of inclusion I have explored here. In the interest of brevity, I have not included justice towards non-humans in this thesis. However, this does not negate the fact that there may be even further important problems of exclusion inherent to social contract theory when such groups are considered.

Hobbesian contractarianism, by and large, bites the bullet on this issue of exclusion, allowing for extensive discrimination, and even unregulated exploitation of certain parties not protected by political contract. Accepting exclusion, discrimination, and exploitation in this way negates much of social contract theory’s appeal, which lies in the ideal of relative formal equality and universal consent of the governed (see introduction to Part III).

Rawlsian theories, however, do not take this avenue. In the case of differences of relative power, Rawlsian theories are adequately able to be inclusive because of their veil of ignorance (Part III, Chapter I). However, even this less self-interested version of social contract theory finds itself unable to provide sufficiently for those who are rationally unable to contract (the severely mentally disabled) and those outside of the social contract (non-citizens and citizens of other nations) without appealing to an independent and preponderant sense of political morality and justice. Without any greater principles overruling self-interest, we may not be able to form a social contract that can be meaningfully called “just.” The fact that even a less self-interested social contract theory is limited in its ability to account for a just social order does not make it a totally irrelevant or debunked theory. However, in admitting that it is in fact unable to account for the entirety of political justice, we must accept that a greater theory of politics,
outside of what contractarianism can conceptually provide, will be necessary before a just social order can be established, even theoretically.
Bibliography


